

**ORGANIZAÇÃO E REPRESENTAÇÃO
DO CONHECIMENTO**
TECNOLOGIAS, SABERES E PRÁTICAS



UNIVERSIDADE FEDERAL DO MARANHÃO

Prof. Dr. Natalino Salgado Filho

Reitor

Prof. Dr. Marcos Fábio Belo Matos

Vice-Reitor



EDITORA DA UNIVERSIDADE FEDERAL DO MARANHÃO

Prof. Dr. Sanatiel de Jesus Pereira

Diretor

CONSELHO EDITORIAL

Prof. Dr. Antônio Alexandre Isídio Cardoso

Prof. Dr. Elídio Armando Exposto Guarçoni

Prof. Dr. André da Silva Freires

Prof. Dr. José Dino Costa Cavalcante

Prof^a. Dra. Diana Rocha da Silva

Prof^a. Dra. Gisélia Brito dos Santos

Prof. Dr. Edson Ferreira da Costa

Prof. Dr. Marcos Nicolau Santos da Silva

Prof. Dr. Carlos Delano Rodrigues

Prof^a. Dr. Felipe Barbosa Ribeiro

Prof. Dr. João Batista Garcia

Prof. Dr. Flávio Luiz de Castro Freitas

Bibliotecária Dra. Suênia Oliveira Mendes

Prof. Dr. José Ribamar Ferreira Junior



ASSOCIAÇÃO BRAISLEIRA DAS EDITORAS UNIVERSITÁRIAS

Valdirene Pereira da Conceição
(Organizadora)

**ORGANIZAÇÃO E REPRESENTAÇÃO
DO CONHECIMENTO**
TECNOLOGIAS, SABERES E PRÁTICAS

São Luís



EDUFMA

2023

Copyright © 2023 by EDUFMA

Revisão: Vera Lúcia G. Sousa

Projeto Gráfico, Capa e Editoração Eletrônica: Amaury Araujo Santos



Dados Internacionais de Catalogação na Publicação (CIP)
Biblioteca Central da Universidade Federal do Maranhão

Organização e representação do conhecimento: tecnologias, saberes e práticas / Valdirene Pereira da Conceição [Org.] — São Luís: EDUFMA, 2023.

362 p. il.

ISBN 978-65-5363-236-3 (E-book)

1. Organização do Conhecimento 2. Tecnologia e Representação do Conhecimento 3. Representação descritiva 4. Indexação 5. Catalogação . I. Conceição, Valdirene Pereira da.

CDD 3020.001

CDU 02:001.82

Marcia Cristina da Cruz Pereira - Bibliotecária / CRB 418 | UFMA-BC/SIB

Criado no Brasil [2023]

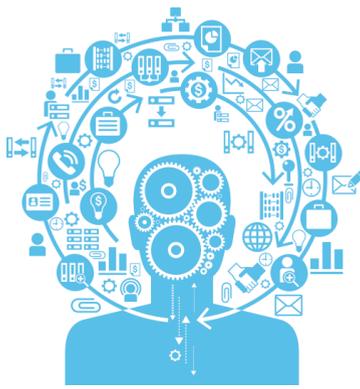
Todos os direitos reservados. Nenhuma parte deste livro pode ser reproduzida, armazenada em um sistema de recuperação ou transmitida de qualquer forma ou por qualquer meio, eletrônico, mecânico, fotocópia, microimagem, gravação ou outro, sem permissão do autor.

EDUFMA | Editora da UFMA | Av. dos Portugueses, 1966 – Vila Bacanga CEP: 65080-805
São Luís | MA | Brasil Telefone: (98) 3272-8157 | www.edufma.ufma.br | edufma@ufma.br

SUMÁRIO

	APRESENTAÇÃO	7
	<i>Valdirene Pereira da Conceição</i>	
	PREFÁCIO	11
	<i>Silvana Maria de Jesus Vetter</i>	
1	SUMARIZAÇÃO DE TEXTOS E REPRESENTAÇÃO DA INFORMAÇÃO NA WEB COM USO DE APIS: aplicação experimental	15
	<i>Janailton Lopes Sousa</i>	
2	A FOLKSONOMIA COMO TÉCNICA DE REPRESENTAÇÃO DA INFORMAÇÃO: estudo das práticas dos usuários da seção banco de cultura do Overmundo	49
	<i>Livia Regina Passos Silva</i>	
3	UTILIZAÇÃO DE ONTOLOGIAS NA ORGANIZAÇÃO DA INFORMAÇÃO NAS BIBLIOTECAS UNIVERSITÁRIAS FEDERAIS DAS REGIÕES NORTE E NORDESTE DO BRASIL	95
	<i>Sandra Gleice da Silva e Silva</i>	
4	EXTRAÇÃO AUTOMÁTICA DE TERMOS: as bases para a estruturação semântica do Vocabulário do Domínio Musical.	149
	<i>Juliana Rabelo do Carmo</i>	

- 5 **DESVENDANDO A DIMENSÃO COGNITIVA DA LEITURA NO PROCESSO DE INDEXAÇÃO**
Tatiane da Silva Mendonca..... 205
- 6 **O PROCESSO DE CLASSIFICAÇÃO BIBLIOGRÁFICA NA ÁREA JURÍDICA: análise comparativa da CDU e CDDir nos domínios do Direito Administrativo e Constitucional** 233
Cintia da Luz Santos
- 7 **IMPACTOS DA POLÍTICA DE INDEXAÇÃO NA RECUPERAÇÃO DA INFORMAÇÃO NAS BIBLIOTECAS UNIVERSITÁRIAS DE SÃO LUÍS**..... 275
Thayland Mafra Muniz
- 8 **A CATALOGAÇÃO DESCRITIVA E SEUS REFLEXOS NAS PRODUÇÕES DAS REGIÕES NORTE E NORDESTE DO BRASIL: dos antecedentes da catalogação aos modelos de conceituais de representação da informação**..... 295
Mauricio José Moraes Costa
- 9 **CATALOGAÇÃO COOPERATIVA NO SISTEMA DE BIBLIOTECAS DA ELETROBRÁS ELETRONORTE** 337
Jousiane Leite Lima



APRESENTAÇÃO

A coletânea **ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO: tecnologias, saberes e práticas** traduz o esforço em reunir e divulgar os estudos e pesquisas desenvolvidos pelo Grupo de Pesquisa e Extensão em “Organização do Conhecimento em Ambientes Culturais (GPEOCAC), do Departamento de Biblioteconomia da Universidade Federal do Maranhão, resultantes das muitas facetas(ensino, pesquisa e extensão) que nos unem, docentes e alunos, nesse universo de que trata a Biblioteconomia e que faz parte do percurso da minha vida acadêmica, nesta Instituição de Ensino Superior – a UFMA.

O ponto em comum, que conecta os textos e tece a rede de relações que se estabelecem entre as diferentes perspectivas de investigação da Biblioteconomia apresentadas em cada estudo, é a Organização do Conhecimento (OC), subcampo da Ciência da Informação e eixo estruturante da formação e prática do bibliotecário. São, portanto, pesquisas acerca de abordagens teórica, aplicada e social dos processos de representação descritiva e temática da informação e organização do conhecimento em diversos ambientes informacionais. Por meio do tratamento da informação, registrada nos mais variados suportes, o conhecimento é divulgado e circula no mundo onde as ideias são absorvidas, analisadas, processadas e julgadas. Logo, o tratamento da informação, indiscutivelmente, é fato político, instrumento de desenvolvimento e fonte de poder.

A Organização do Conhecimento, por meio de seu instrumental teórico- metodológico e de recursos mecanizados de processamento para acesso, uso e reuso da informação, propicia a ordem na enorme coleção de dados, no contexto das mídias sociais, da cibercultura, da Ciência aberta e de *linked data*. Essas abordagens aparecem transversalmente nas pesquisas aqui apresentadas, sob as perspectivas das tecnologias, dos instrumentos semânticos de representação, da informação e da dimensão cognitiva dos saberes e práticas de organização do conhecimento.

A título de exemplo, o texto de abertura da coletânea, de autoria de Janailton Lopes Sousa, trata da aplicação experimental da sumarização de textos e representação da informação na web com o uso de API (do inglês *Application Programming Interface*). O estudo da Folksonomia, como técnica de representação da informação, realizado pelos usuários da seção banco de cultura do Overmundo, é apresentado por Livia Regina Passos Silva. A coletânea traz ainda a pesquisa que tem como título “Utilização de Ontologias na Organização da Informação nas Bibliotecas Universitárias Federais das Regiões Norte e Nordeste do Brasil”, realizada por Sandra Gleice da Silva e Silva.

O texto “Extração automática de termos: bases para a estruturação semântica do Vocabulário do Domínio Musical,” resultado do trabalho realizado por Juliana Rabelo do Carmo, põe em evidência a intersecção entre Linguística Computacional, Processamento Língua Natural (PLN) e Organização do Conhecimento e fundamenta-se na pesquisa aplicada e utilização do arcabouço teórico-metodológico do PLN na elaboração de vocabulário controlado em domínios especializados.

Outros textos também compõem esta publicação e tratam dos meandros do processo de indexação, a exemplo da investigação intitulada “Desvendando a dimensão cognitiva da leitura no processo de indexação”, de Tatiane da Silva Mendonça, que descreve as estruturas cognitivas (simples – memória e imaginação e complexas – compreensão e raciocínio) utilizadas na determinação do assunto e indica um roteiro de leitura técnica, para identificação, seleção e determinação dos conceitos que irão representar o tema tratado no documento. A autora observa que a leitura, de um modo geral, trata-se de uma combinação de fatores, sejam eles linguísticos, extra-linguísticos e cognitivos, e, como tal, uma vez combinados, são reveladores do sentido que circula no texto.

“O Processo de Classificação Bibliográfica na Área Jurídica: análise comparativa da CDU e CDDir nos domínios do Direito Administrativo e Constitucional”, de Cintia da Luz Santos, é um texto que indica pontos de análise do distanciamento e aproximação entre os dois sistemas de classificação do conhecimento. O texto apresenta ainda abordagens correlatas sobre o bibliotecário e sua ação num determinado marco histórico; o reconhecimento de processos de exclusão, diferenciação e dominação como historicamente construídos; e a compreensão contemporânea da esfera social e cultural da Organização do Conhecimento.

O texto intitulado “Impactos da política de indexação na recuperação da informação nas bibliotecas universitárias de São Luís,” escrito por Thayland Mafra Muniz, aponta os elementos estruturantes de uma políti-

ca de indexação, como ferramenta auxiliar na tomada de decisão e diretriz do processo de indexação.

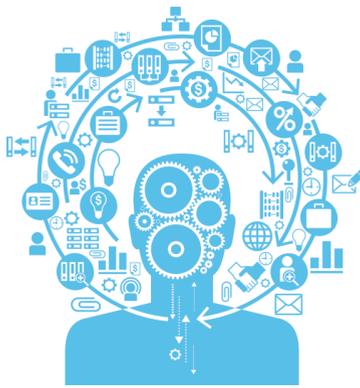
Outros temas de relevante interesse para os estudos em Biblioteconomia são abordados no texto “Catalogação Descritiva e seus Reflexos nas Produções Técnicas das Regiões Norte e Nordeste do País: dos antecedentes da catalogação aos modelos conceituais de representação da informação”, em razão disso também são contemplados nesta publicação. A autoria é de Mauricio José Moraes Costa, cuja pesquisa teve como objetivo identificar as tendências das temáticas e limitações nos estudos em catalogação nas regiões norte e nordeste nos Programas de Pós-Graduação em Biblioteconomia e Ciência da informação.

Partindo-se, nesta coletânea, de textos com um embasamento mais teórico e conceitual, finalizamos com a apresentação de um relato de experiência sobre a “Catalogação Cooperativa no Sistema de Biblioteca da Eletrobrás Eletronorte”, de Jousiane Leite Lima, o qual traz uma reflexão acerca da concepção de sistema de catalogação cooperativa do bibliotecário e de suas práticas, da estrutura técnica e tecnológica necessária para a realização desse serviço, assim como as vantagens para a biblioteca no que concerne à racionalização de tempo e recursos.

Esperamos com esta primeira coletânea do GPEOCAC contribuir para as pesquisas no campo da Organização do Conhecimento, principalmente aquelas interessadas em analisar a prática do bibliotecário e sua participação como sujeito nos processos de representação da informação, de vinculação social ou mesmo de mediação e valorização de tratamento da informação e do conhecimento em bibliotecas, arquivos, museus, galerias e centros culturais contemporâneos.

São Luís(MA), julho de 2022

Valdirene Pereira da Conceição



PREFÁCIO

O desenvolvimento da sociedade, atualmente, exige foco no acesso à informação e ao conhecimento, bem como, na capacidade de aprender, inovar e desenvolver competências, considerando que o novo instrumental tecnológico viabiliza fluxos de informação em grande volume. O aumento crescente no volume de documentos, especialmente nas unidades de informação, nos insere na busca por ferramentas técnicas e tecnológicas, e por um arcabouço teórico que nos auxilie no processo de organização e representação do conhecimento.

Nessa perspectiva, a obra **ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO TECNOLOGIAS: saberes e práticas**, organizada por Valdirene Pereira da Conceição, se constitui em um desses achados que nos conduz com maestria ao mundo da organização e representação do conhecimento. Professora e Pesquisadora vinculada ao Departamento de Biblioteconomia/UFMA, Valdirene coordena um grupo de pesquisa que vem procurando desenvolver estudos com temas originais e metodologias inovadoras, na área da organização e representação do conhecimento. Sua vasta experiência em ensino e pesquisa, e o seu compromisso em realizar produções científicas que sejam úteis à reflexão e ao desenvolvimento de competências bibliotecárias, para lidar com a complexa tarefa de organização da massa documental existente, são fatores que culminaram na realização desta excelente obra, em colaboração com egressos do Curso de Biblioteconomia/UFMA. Este livro apresenta estudos que procuram desvendar os mistérios que permeiam a área da representação do conhecimento, e buscar soluções para superação dos desafios impostos pela sociedade mediada por Tecnologias Digitais de Informação e Comunicação.

No primeiro capítulo *“Sumarização de texto e Representação da informação na web com uso de APIs: aplicação experimental”* de Janailton Lopes de Sousa, a ênfase recai sobre a sumarização de textos, atividade que

procura identificar informações que contribuam na produção de textos a partir de textos, e podem servir de indexadores, ou seja, permite uma descrição compacta e fiel do documento. O autor destaca a sumarização de textos com o uso do Processamento da Linguagem Natural (PLN), e algumas formas de realizá-la, trazendo como exemplo a elaboração de mapas conceituais e mentais.

No segundo capítulo *“Folksonomia como técnica de representação da informação: estudo das práticas dos usuários da seção banco de cultura do Overmundo”*, Livia Regina Passos Silva discute sobre a folksonomia no contexto da web 2.0, e ao analisar os critérios de indexação/etiquetagem da seção banco de produtos culturais digitais do site Overmundo, que é um website colaborativo voltado para a cultura brasileira verificou que a Folksonomia se enquadra nos esquemas de representação, por se tratar de ferramenta semântica, que, embora não tenha vocabulário definindo, facilita a identificação de objetos em meio digital.

No terceiro capítulo *“Utilização de ontologias na organização da informação nas bibliotecas universitárias das Regiões Norte e Nordeste do Brasil”*, Sandra Gleice da Silva e Silva aponta a Web Semântica como ferramenta que acrescenta significado aos documentos disponíveis em rede, e a ontologia como fundamental na representação, compartilhamento, reuso e manutenção do conhecimento, por se tratar de língua franca da Web Semântica. Enfatiza as ontologias como estruturas de representação do conhecimento na web, dando destaque às técnicas para desenvolvê-las, implementá-las, e usá-las na representação do conhecimento.

No quarto capítulo *“Extração automática de termos: as bases para a estruturação semântica do vocabulário controlado”*, Juliana Rabelo do Carmo aponta que a análise documental de acervos especializados, como as fontes de informação em música, tem sido precária. Ela analisou o cenário prático-conceitual da indexação e representação da informação, para estruturar um vocabulário de domínio musical, analisar os léxicos produzidos nesta área e relacioná-los a outros conceitos, com uso de categorização. Para a extração automática dos léxicos, utilizou o E-terminos, e obteve 922 itens lexicais, entre unigramas, bigramas e trigramas que podem aprimorar a construção de vocabulários fundamentados em PLN.

No quinto capítulo, *“Desvendando a dimensão cognitiva da leitura no processo de indexação”*, Tatiane da Silva Mendonça analisou os aspectos cognitivos que auxiliam o bibliotecário no processo de leitura realizado na indexação. Ela constatou que na visão dos bibliotecários, a indexação é um processo realizado por meio de leitura técnica do documento para extração de termos mais representativas de seu conteúdo e os instru-

mentos e métodos utilizados foram: vocabulário livre, AACR2, tabela de Cutter, buscadores (Google), catálogos ou índices, tesouros e CDU, mostrando que os bibliotecários tendem a seguir os manuais de catalogação e indexação, e que a leitura documentária é uma atividade complexa que vai além da decodificação de símbolos e sinais gráficos e visuais presentes no textos e documentos.

No sexto capítulo, *“O processo de classificação bibliográfica na área jurídica: análise comparativa da CDU e CDDir nos domínios do Direito Administrativo e Constitucional”*, Cíntia da Luz Santos ressalta que a área Jurídica tem sua própria classificação para distinguir seus diversos ramos (administrativo, constitucional, civil, tributário, processual), que é a Classificação de Direito (CDDir). Ela descreveu e analisou a representação temática do domínio jurídico Direito Administrativo e Constitucional, adotados pela Biblioteca do TJ-MA e pela Biblioteca Central da UFMA, visando compreender a atividade da Classificação Bibliográfica na área jurídica, estruturou conceitos na CDU e CDDir, e as comparou: enquanto a CDU abrange todas as áreas do conhecimento, a CDDir é exclusiva da área jurídica.

No sétimo capítulo *“Impactos da política de indexação na recuperação da informação nas bibliotecas universitárias de São Luís”*, Thayland Mafra Muniz discute sobre a política de indexação como diretriz para garantir padrões de qualidade e excelência no processo de representação do conhecimento. Logo, não é lista de procedimentos, mas uma filosofia que define rotinas e representa os anseios e os objetivos da biblioteca, passando por todo o processo do ciclo documentário. Ela apresenta a política de indexação nas bibliotecas universitárias de São Luís destacando como seus principais objetivos atender às necessidades de informação de professores, alunos e da comunidade por facilitação do acesso, consulta, e recuperação da informação e apoio às atividades de ensino, pesquisa e extensão da instituição.

No oitavo capítulo *“A catalogação descritiva e seus reflexos nas produções das regiões Norte e Nordeste do Brasil: dos antecedentes da catalogação aos modelos conceituais de representação da informação”*, Maurício José Moraes Costa ressalta o papel da catalogação como geradora do Tratamento Descritivo da Informação (TDI) associada à sua usabilidade, aos metadados, ontologias, modelagens de dados e conteúdos. Enfatiza os modelos conceituais de representação da informação e apresenta o RDA como norma que se alinha aos modelos conceituais e converge recursos com Marc 21, Dublin Core e outros, devido à sua flexibilidade e capacidade de extensão, no processo de descrição técnica de recursos

e conteúdos. Trata da catalogação descritiva refletida nas produções do Norte e Nordeste brasileiro.

No nono capítulo “*Catalogação cooperativa nos sistemas de bibliotecas da Eletrobras/Eletronorte*”, Jousiane Leite Lima, procurou conhecer a aplicabilidade da catalogação cooperativa nas bibliotecas da Eletrobras Eletronorte com o uso do SophiA Biblioteca. Constatou que os bibliotecários dessa unidade de informação utilizam a catalogação cooperativa em suas atividades e essa ação foi facilitada pela automação das bibliotecas e criação de padrões de intercâmbio de dados. Constatou também que fazem uso do sistema SophiA Biblioteca, por atender suas necessidades da unidade de informação.

Os temas abordados neste livro reúnem contribuições importantes sobre os aspectos teóricos, conceituais e práticos da área de organização e representação do conhecimento. Compõe, portanto, o rol de estudos na área da Ciência da Informação que vêm suprir a lacuna de pesquisas que se preocupam em encontrar soluções para a complexa tarefa de organizar e representar o conhecimento, na sociedade mediada por novas tecnologias.

Silvana Maria de Jesus Vetter



SUMARIZAÇÃO DE TEXTOS E REPRESENTAÇÃO DA INFORMAÇÃO NA WEB COM USO DE APIS: aplicação experimental

Janailton Lopes Sousa

INTRODUÇÃO

O pós-guerra marca o grande investimento em tecnologias, as de informação e comunicação ganham destaque, pois almejavam uma conexão em nível global, o progresso tecnológico abarcaria inúmeros países num ritmo latente, onde novas descobertas seriam constantes. Esse progresso tecnológico acelerado, ganha mecanismos como o computador que por sua capacidade de processamento e armazenamento modifica a produção de informação e propagação do conhecimento.

Nota-se a permeabilidade das novas tecnologias nas vidas das pessoas, alterando as relações de tempo e espaço, aumentando a capacidade de pensar e a inteligência humana. Verifica-se deste modo, sua grande influência, devido ao gigantesco processamento de informação. Entre os anos de 1970 e 1980 o uso de computadores ganha intensidade, mas é a partir da década de 90 que a comunicação em rede eclode, potencializando uma maior interação de quem utiliza esse suporte informacional.

Essas tecnologias provocam interferências diretas nas vidas das pessoas, pois são tecnologias de inteligência que amplia exponencialmente a capacidade de tratar informações e transformá-las em conhecimento. Silveira (2005) ressalta ainda que o exercício da leitura emula a criatividade e fomenta os processos de reflexões cada vez mais complexas, elevando assim a capacidade de pensar. Expondo que precisamos nos adaptar a esse novo contexto que de certa forma exclui quem não consegue desenvolver competências necessárias para usufruir de suas potencialidades.

Neste sentido o surgimento da internet tem considerável influência nas leituras em um aparelho que permita uma apresentação gráfica em uma tela de computador, comportando assim a abertura de novos rumos para a leitura em tela. A internet inseriu um processo de avanço sem precedentes por meio dos navegadores que permitem um amplo acesso a diversas informações em diferentes formatos delineando uma marcha incessante para o progresso.

As Tecnologias Digitais de Informação e Comunicação (TDICs) vieram impulsionar o processo de comunicação cotidiana, e, com os novos recursos informacionais, que permitem um acesso rápido e volumoso às informações em nível global e local. Por isso, torna-se necessário estudar suas vantagens e desvantagens, diante do impacto que gera na sociedade, desde o surgimento destes novos suportes, à necessidade dos profissionais da informação para se adaptarem a essas novas tecnologias são exigidas como forma de aperfeiçoamento em sua área de atuação.

Fundamentado neste contexto tecnológico e no aporte teórico que dão embasamento na área de Ciência da Informação especificamente na Biblioteconomia, buscou-se relacionar ambos para concepção de algo, que pudesse agregar conhecimento a esta área. Para isto recorreu-se às contribuições teóricas da Comunicação, Ciência da Computação, Filosofia e Psicologia que forneceram subsídios necessários para concepção deste trabalho. As Tecnologias Digitais de Informação e Comunicação (TDICs) vieram impulsionar o processo de comunicação cotidiana, e, com os novos recursos informacionais, que permitem um acesso rápido e volumoso às informações em nível global e local. Por isso, torna-se necessário este tipo de estudo diante do impacto que gera na sociedade, desde o surgimento destes novos suportes, à necessidade dos profissionais da informação para se adaptarem a essas novas tecnologias são exigidas como forma de aperfeiçoamento em sua área de atuação.

INDEXAÇÃO E ANÁLISE DE CONTEÚDO

O uso de teorias de organização e classificação do conhecimento dá suporte a uma etapa posterior a classificação de um documento portador de um determinado conhecimento, sendo a classificação uma etapa de categorização conceitual, a indexação prevê uma tradução para a representação desse conceito por meio de termos extraídos do documento. Lancaster (2004), esclarece que esta confusão é muito comum entre bibliotecários quando utilizam a indexação como mecanismo de classificação, isso por conta de diferentes terminologias. Para Guinchat; Menou (1994, p.167):

A classificação é uma operação da descrição de conteúdo de documentos, pela qual determina-se o assunto principal e, eventualmente, um ou dois assuntos secundários que são traduzidos para o termo mais apropriado da linguagem documental utilizada. Os termos da classificação são expressos, em geral, por notações. Mas podem ser utilizados também termos genéricos de uma linguagem combinatória, como os nomes dos campos ou os grupos de uma lista de descritores ou de um tesouro.

Neste sentido não será destacada a problemática terminológica durante a etapa de classificação, mas o processo teórico que envolve a indexação por meio de autores clássicos desta área como Lancaster (2004); Guinchat e Menou (1994), e da análise de conteúdo como Bardin (1977), buscando suas relações conceituais, afim de ampliar essa discussão. Uma vez que a classificação por meio de notação numérica é empregada frequentemente por bibliotecários por meio da análise de assunto. Já a classificação por meio de descritores ou de um tesouro envolve um aspecto conceitual mais marcante, tendo em vista que o conceito é a unidade básica do conhecimento.

A classificação bibliográfica da *Dewey Decimal Classification* e *Classificação Decimal Universal*, por exemplo, busca prioritariamente representar um determinado documento por meio de uma notação atribuída a determinado assunto de uma área do conhecimento. Entretanto existem outras formas de se representar um documento, seja pelas referências, descrição bibliográfica, indexação, resumos, cabeçalhos de assunto, sumários, índices catálogos entre outros. Formas estas que remetem sempre ao documento original.

A indexação compõe uma dessas formas, que comumente é empregada no ambiente do bibliotecário que utiliza de técnicas de análise de assunto e de conteúdo para elaborar uma forma coesa de representação documental. Apesar de comum necessita de especial atenção quanto às técnicas de análise conceitual, temática e de conteúdo do documento visando um bom aproveitamento no processo de recuperação da informação.

A Associação Brasileira de Normas Técnicas (ABNT) define a indexação como ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação (ABNT, 1992). Uso de técnicas e padrões internacionais e nacionais na indexação objetiva equidade nos processos de representação dos conteúdos documentais.

Segundo Guinchat e Menou (1994), a indexação é uma das formas de descrição de conteúdo. É a operação pela qual escolhe-se os termos

mais apropriados para descrever o conteúdo de um documento. Esta etapa intelectual pode ser feita por uma ou várias pessoas sobre o assunto do documento a ser processado, a extração de termos, descritores e a linguagem documental como um vocabulário controlado é um ponto de grande importância nesta etapa, para Santos (2011, p.7):

A indexação é a operação que consiste em recuperar, selecionar e exprimir por meio de termos pertencentes a uma ou várias linguagens documentais as informações contidas nos documentos. Trata-se de uma operação de descrição interna, cujo objeto é o conteúdo intelectual dos documentos.

No aspecto manual ou semiautomático exige do bibliotecário conhecimento prévio sobre determinado assunto a fim de extrair termos significativos que possam representar determinado conteúdo de um documento, por esta razão:

O processo de indexação por assuntos, dentro das técnicas documentais, não pode contar com normas que verdadeiramente apoiem a objectividade da análise do documento, determinem, com precisão, a decisão correcta a tomar perante as opções colocadas pela representação dos conceitos, de forma a garantir a sua coerência e conferir uniformidade aos instrumentos de pesquisa disponibilizados ao universo utilizador. (MENDES; SIMÕES, 2002, p.13).

Rubi (2009) afirma que a representação desse conteúdo é realizada por meio de conceitos. Esses conceitos, por sua vez, serão representados ou traduzidos em termos advindos de uma linguagem documentária. Equilibrar essa relação entre a subjetividade do indexador e as estruturas padronizadas de uma linguagem documentária como um vocabulário controlado, por exemplo, deve ser um objetivo básico do indexador.

De acordo com Leiva (2012), a indexação gera palavras-chave, índices ou os cabeçalhos de assunto de um documento. Para obtê-los, previamente foi desencadeada uma sucessão interativa e simultânea dos processos mentais. Isto portanto, pode diversificar-se de acordo com o indexador. Entretanto deve-se observar que este processo direciona-se a objetividade, precisão e recuperação da informação. A relação entre o processo cognitivo e a indexação do conteúdo de um documento reflete diretamente na eficiência da recuperação do mesmo.

A indexação é o processo intelectual que envolve atividades cognitivas na compreensão do texto e a composição da representação do documento, (LIMA, 2003). Este processo cognitivo envolve a interpretação, análise e desconstrução de um documento e sua tradução para uma linguagem documentária que forneça elemento que possam representar aquele documento.

O processo de indexação pode ser simplificado em três etapas que consistem na análise conceitual que se preocupa em identificar o assunto do documento por meio de seu conteúdo, na tradução que consiste na substituição de uma linguagem natural para uma linguagem documentária e a representação que é a nova versão do documento cujos atributos documentais representam o documento original. Segundo Rubi (2009), autores como Unisit, (1981); Chaumier, (1986); Lancaster, (2004); ABNT 12.676, (1992); Van Slype, (1991) e Robredo, (2005) divergem quanto ao número de etapas da indexação, porém, elas tratam basicamente das mesmas operações. São elas:

- a) Análise: leitura e segmentação do texto para identificação e seleção de conceitos;
- b) Síntese: construção do texto documentário com os conceitos selecionados. Está relacionada especificamente à elaboração de resumos;
- c) Representação: por meio de linguagens documentárias.

O Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde, mais conhecido como BIREME (2008), concebe a indexação como o processo pelo qual é descrito o conteúdo de um documento mediante descritores, depois da sua leitura técnica e análise. Independente dos autores é possível afirmar que análise é uma etapa inicial que envolve um processo cognitivo do indexador que utiliza técnicas e métodos que buscam objetivar este processo. A etapa da análise do conteúdo do documento é um importante momento onde se extrai os conceitos contidos no documento por meio de seus descritores. Para Bardin (1977, p.34):

A análise de conteúdo pode ser uma análise dos <<significados>> (exemplo: a análise temática), embora possa ser também uma análise dos <<significantes>> (análise léxica, análise dos procedimentos). Por outro lado, o tratamento descritivo constitui um primeiro tempo do procedimento, mas não é exclusivo da análise de conteúdo. Outras disciplinas que se debruçam sobre a linguagem ou sobre a informação, também são descritivas: a linguística, a semântica, a documentação.

Ressalta que a análise de conteúdo surge como um conjunto de técnicas de análise das comunicações, que utilizam procedimentos sistemáticos e objetivos de descrição de conteúdo das mensagens. Apesar de muito comum nas ciências sociais, esta técnica pode ser utilizada por qualquer um em qualquer área do conhecimento e de forma interdisciplinar estas técnicas de análise dialogam especialmente com a Biblioteconomia, no que diz respeito aos elementos de análise e descrição de conteúdo dos documentos. Bardin (1977), organiza a análise de conteúdo basicamente em três fases:

- a) A pré-análise é a fase em que se organiza o material a ser analisado com o objetivo de torná-lo operacional, sistematizando as ideias iniciais.
- b) A segunda fase consiste na exploração do material com a definição de categorias e a identificação das unidades de registro e das unidades de contexto nos documentos.
- c) A terceira fase diz respeito ao tratamento dos resultados, inferência e interpretação.

Diante da abrangência que a análise de conteúdo compreende, seu uso pode ser empregado de acordo com especificidade do que será analisado, neste caso ressalta-se sua contribuição na indexação, para que se eliminem ao máximo as representações que não se adequam aos conteúdos primários dos documentos.

De acordo com Lancaster (2004); Guinchat; Menou (1994) e Rubi (2009) a indexação obedece a alguns princípios básicos que devem ser destacados, são eles: exaustividade, que exige o maior número possível de termos e especificidade, onde um número de termos sobre determinado assunto seria empregado com maior extensão, além da capacidade de revocação que se relaciona ao número de documentos recuperados e a precisão que se identifica com o número de documentos recuperados pela relevância ou interesse da solicitação.

A indexação é um elemento de valiosa importância dentro de um Sistema de Recuperação da Informação – SRI, nesta seção são destacados apenas os pontos conceituais e sua aplicação manual, tendo portanto uma participação significativa dos processos cognitivos do bibliotecário, entretanto este processo também pode ser semiautomático ou automático.

O uso de tecnologias na indexação

Os Sistemas de Recuperação da Informação (SRI) sempre demonstraram sua eficiência no gerenciamento e acesso a documentos impressos nas unidades de informação. Já o indexador valendo-se ao máximo de sua memória, capacidade de raciocínio, associação, leitura, interpretação e extração de descritores que representem determinado documento, destaca portanto a subjetividade do processo de indexação manual e inserção de descritores que comprometam a eficiência de um SRI.

Isto foi modificado ao longo do tempo com a inserção de tecnologias digitais que alteraram as formas de indexar conteúdo de documentos, mesmo não sendo algo novo, pois Gomes (1989), destaca que desde

a década de 50 o *KeyWord In Context* – (KWIC) direcionava o uso de tecnologias no processo de indexação e desta forma chama a atenção do bibliotecário para atentar-se ao uso de tecnologias no contexto informacional que começava a emergir.

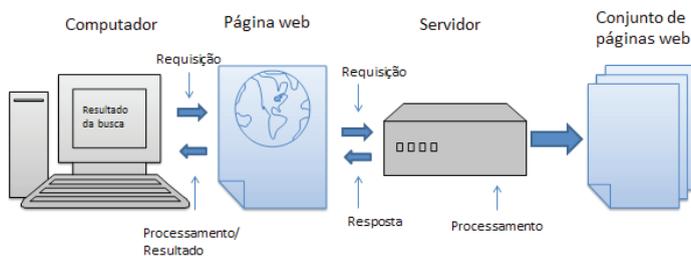
O problema da indexação manual segundo Gomes (1989), é que não existe uma única leitura do documento e como consequência, a indexação de qualquer documento pode variar de indexador para indexador e no mesmo indexador, no tempo. Estas variações oriundas das subjetividades dos indexadores modelam as leituras que estes têm do mundo e dos documentos que passam por suas mãos, entretanto isto pode ser amenizado com uso de técnicas de leitura para se alcançar a objetividade na indexação.

A indexação pode ser manual quando não se utiliza nenhuma tecnologia digital para se extrair os termos contidos nos textos e, caso estes não se adequem o indexador pode atribuir termos que melhor representem o documento, recorrendo, portanto a sua memória, leituras e relações conceituais formuladas pelo próprio indexador.

A indexação pode ser semiautomática quando se utiliza um computador para operar softwares que possam extrair termos a partir de textos, formando um *corpus* que será posteriormente analisado pelo indexador. É automática quando utiliza um mecanismo de extração, análise e representação do documento sem intervenção humana.

É importante destacar que a indexação automática no ambiente computacional é bem comum, pois o que os navegadores web (*browsers*) basicamente fazem é isto, indexam páginas web a partir de palavras utilizadas no campo de busca e que estejam contidas nos URLs das páginas ou na descrição do site, conforme apresentado na Figura 1.

Figura 1 – Representação de busca de uma página web



Fonte: Sousa, (2016)

Todo e qualquer tipo de conteúdo pesquisável está contido em algum lugar, seja no modelo tradicional ou tecnologicamente moderno dos SRI. A partir deste armazenamento que a *priori* são constituídos de dados é possível organizá-los e gerenciá-los com auxílio de softwares específicos.

Na Biblioteconomia quando se fala em base de dados pensa-se normalmente em alimentá-la manualmente, preenchendo diversos campos que compõe uma tabela de dados que se referem a um documento a ser inserido dentro de sistema de automação em Unidades de Informação (U.I).

Entretanto, tecnologias emergentes da Inteligência Artificial (IA) como aprendizado de máquina (*Learning Machine*) e Processamento da Linguagem Natural (PLN) consistem basicamente em tornar a ação humana compreensível por uma máquina, para se automatizar tarefas humanas.

A indexação automatizada consiste em fazer o computador reconhecer palavras que aparecem no título, no resumo do documento, ou no seu próprio texto. Os termos reconhecidos são incorporados em um arquivo de pesquisa e servem para recuperar o documento. (GUINCHAT; MENOU, 1994, p. 182).

As palavras representadas no formato binário normalmente estão armazenadas em bancos de dados que enviam repostas de solicitação proveniente de um Sistemas de Gerenciamento de Banco de Dados (SGBD). O uso de descritores em um sistema a ser utilizado dentro de uma U.I. implica na prática, a precisão e satisfação de quem opera uma busca por determinado assunto de acordo com os termos indexados e armazenados no banco de dados por meio de um software utilizado. Portanto quando se fala em indexação automática pressupõe-se que os descritores que representam um documento já foram extraídos, armazenados numa base de dados e elencados dentro de um resultado de busca.

De acordo com Cendón (2002), as bases de dados são especialmente adequadas para responder a perguntas multifacetadas, porque muitos, se não todos os campos dos seus registros, são indexados e, portanto, pesquisáveis. Existem inúmeras possibilidades de uso e análise, redirecionando aos grandes geradores de dados (pessoas e organizações públicas e privadas) a possibilidade de sempre extrair informações que podem vir a se tornar um conhecimento.

As bases de dados bibliográficas estão povoadas de documentos textuais cuja alimentação depende de profissionais que indexam esses textos popularizando descritores extraídos ou atribuídos por eles. Ape-

sar disso existem softwares que fazem extração automática de termos. Atualmente o mais próximo do correto quando se trata de assuntos relacionados à tecnologia na indexação semiautomática seria a extração de candidatos a termos ou extração automática; termos estes que depois de extraídos poderão ser utilizados para auxiliar no processo de indexação, construção de vocabulários controlados, de um tesouro, entre outros.

Apropriar-se de conhecimentos sobre tecnologias digitais é tarefa do profissional da informação, ainda que não domine completamente tais instrumentos tecnológicos. Conhecer é o primeiro passo para integrar dinamicamente com estas tecnologias. Seu uso em Bibliotecas Digitais, Gerenciamento Eletrônico de Documento (GED), armazenamento em nuvem e outros tipos de ambientes virtuais exigem especialmente desse profissional um diálogo verdadeiro com a tecnologia.

Os avanços em mineração de dados e textos, big data, PLN, web semântica e ontologias devem compor o espaço de curiosidade e formação do profissional da informação, que deve acompanhar minimamente estas discussões. Converter estas discussões em aplicações práticas no desempenho deste profissional é um desafio a ser enfrentado. Além disso, os processos de criação e inovação tecnológica devem seguir a prática do bibliotecário, pois a além de consumir ele deve produzir e/ou inovar tecnologias.

O PROCESSAMENTO DA LINGUAGEM NATURAL

Fazer uma máquina entender exatamente o que queremos dizer é um desafio dos estudos desenvolvidos pelo Processamento da Linguagem Natural (PLN). A linguagem faz parte da vida humana e, usualmente utiliza-se a mesma forma natural da fala com outras pessoas, para se efetuar uma busca na web.

Todos os dias são publicados diversos conteúdos expressos na forma de texto que são armazenados em grandes bases de dados. A recuperação da informação neste espaço é feita à medida que as palavras que utilizadas para representar o documento, seja a mesma que uma pessoa comum vai utilizar para localizá-la. O computador não possui o pleno domínio de distinção do conteúdo dos documentos recuperados para saber qual será a resposta certa.

Segundo Vieira; Lopes (2010) o Processamento de Linguagem Natural (PLN) é uma área de Ciência da Computação que estuda o desenvolvimento de programas de computador que analisam, reconhecem e/ou geram textos em linguagens humanas, ou linguagens naturais.

Diariamente são produzidos inúmeros textos virtuais e indexadores da *web* constroem índices, atualizam páginas, endereços e otimizam as buscas. Eles são conhecidos como *crawlers*, *spiders*, *robots*, o mais conhecido sem dúvida é o *googlebot*. A grande questão é que esses indexadores buscam as mesmas palavras requeridas no campo de busca em elementos que compõe ou representam aquela página como URLs, *tags* e descrição da página, não acompanhando a evolução da própria linguagem exceto se alguém publicar algo a respeito do novo assunto.

Estas questões de linguagem interferem diretamente no tipo de informação a ser recuperada, pois a linguagem natural humana é acompanhada de mutações, variações linguísticas, contrações, estrangeirismos, gírias e diversas outras manifestações. Apesar dos avanços em inteligência artificial, o computador se beneficia de uma linguagem formal, enquanto a natural pode conter subjetividades e interpretações diversas.

Apesar disto, é notório que existem diversos softwares que analisam estruturas gramaticais e morfológicas com o intuito de identificar uma ordem semântica em frases e textos para se extrair algum significado disto. Leitão (2006), apresenta três abordagens para o PLN, são elas, simbólica, empírica e neural. A abordagem simbólica é capaz de tratar todos os níveis da análise linguística de uma Linguagem Natural (LN). A abordagem empírica apoia-se no uso de grandes quantidades de dados e procedimentos estatísticos e a abordagem neural fundamenta-se em Redes Neurais Artificiais para implementação de redes semânticas.

Além disso, este processamento pode ser feito por meio de análise morfológica, sintática, semântica ou pragmática. A morfológica consiste em analisar palavras no aspecto de formação e suas características gramaticais. A Sintática (*parsing*) preocupasse com a estrutura das sentenças. A Semântica refere-se ao significado das sentenças. E a Pragmática ocupa-se dos enunciados e relações sintagmáticas para se identificar o contexto das sentenças.

Neste sentido, a Extração Automática de Candidatos a Termos (EACT) caracteriza-se como uma das etapas do PLN. Isto consiste no uso de um software que por meio de um corpus textual se beneficiará de um dos tipos de análise supracitada acima.

A extração automática de termos é um mecanismo utilizado para auxiliar a indexação semiautomática, isto é possível por meio da extração de candidatos a termos. Segundo Zavaglia et al (2005), termos são unidades linguísticas, isto é, palavras ou combinações de palavras, designando conceitos ou entidades de um campo altamente especializado da ativi-

dade humana. Estes candidatos a termos pertencem a um determinado domínio incorporado a uma determinada área do conhecimento.

Essa extração automática por si só não determina quais termos serão utilizados, pois o processamento dos mesmos não os classifica completamente a ponto de determinar exatamente quais termos deverão ser utilizados, por isso o mais correto como supracitado seria extração de candidatos a termos, que por meio de um *software* identifica determinados padrões lexicais que vão desde a estrutura morfológica à composição semântica, identificáveis por meio de termos simples unigrama e compostos, bigramas, trigramas ou mais palavras (n-gramas).

A extração automática de termos é utilizada pelo PLN, para traduzir a linguagem humana para uma linguagem de máquina, posterior a essa tradução o aprendizado de máquina é um mecanismo importante para que, além de identificar esses termos a máquina possa saber o que significa cada um, assim como a sua posição lexical. Isto é basicamente ensinar uma máquina a ler e compreender um texto como um ser humano.

A criação de um *corpus* é outro passo importante, pois o *corpus* constituirá o universo do estudo, este pode ser resultado de materiais que já se encontram no formato digital (estruturado) ou tradicional (não estruturado) o que exige a conversão do mesmo para ser estruturado. Posterior a isto, o formato mais aceitável entre os softwares de extração é o *txt*, um arquivo de texto simples. A análise por meio do software pode ser feita com o método estatístico usando a frequência das palavras (n-gramas), o método linguístico que se fundamenta em expressões linguísticas, indicadores estruturais e padrões morfossintáticos e o método híbrido que beneficia-se de ambos os métodos supracitados. Vieira e Lopes (2010), corroboram, que a extração de termos pode ser feita segundo três abordagens:

- a) A abordagem linguística que geralmente envolve a identificação da organização das frases dos textos (*corpus*) sintaticamente.
- b) A abordagem estatística os candidatos a termos que são extraídos segundo uma análise da sua frequência de ocorrência no *corpus*.
- c) A abordagem híbrida uma combinação das duas outras abordagens, onde tanto conceitos linguísticos, quanto informações estatísticas são consideradas.

A extração automática de termos pode ser aplicada para diversos fins, onde o aprendizado de máquina e interpretação humana convergem para apresentação de resultados, que processam a linguagem natural,

para uma linguagem de máquina e o resultado deste processamento retorna ao ser humano em linguagem compreensível, utilizando técnicas de extração para construção de insumos que possibilitem a recuperação da informação em um SRI aplicado à um determinado domínio. Para Teline (2004, p.16):

A tarefa de extração de termos é crucial para várias aplicações, tais como Recuperação de Informação (RI), Sumarização Automática, Indexação e Classificação de Textos, Alinhamento de Textos Bilingües, Tradução, Recursos Básicos de PLN, Edição Suportada por Computador, Corretores Gramaticais e Geração de Língua Natural (LN).

Estas diversas aplicações da extração automática convergem para a interação da linguagem humana para a linguagem de máquina. Com vistas em aperfeiçoar a o processo de recuperação da informação em ambientes automatizados.

Sumarização de Textos

O PLN envolve diversas aplicações na área da Linguística, Ciência da Computação e Ciência da Informação onde se enquadra a Biblioteconomia, estas aplicações envolvem, sobretudo a sumarização de textos e indexação no processo de recuperação da informação. Pardo (2008), percebe a sumarização como uma subárea de pesquisa do PLN. Álvarez (2007), entende que a sumarização consiste na criação de uma descrição compacta de um documento ou uma coleção de documentos, porém preservando seus significados-chave.

A sumarização preocupa-se com a condensação do conteúdo de documentos considerados textos-fonte, esses textos são processados e por meio deles é extraída uma síntese que pode indicar do que trata o documento, tudo isto de modo automático.

A sumarização textual, como tarefa de produção de textos, obedece, de um modo geral, às mesmas características delineadas acima para a produção de um discurso. Entretanto, ela se distingue por uma restrição fundamental: a de transmitir a mensagem essencial, de forma concisa. (RINO; NUNES, 2005, p.5):

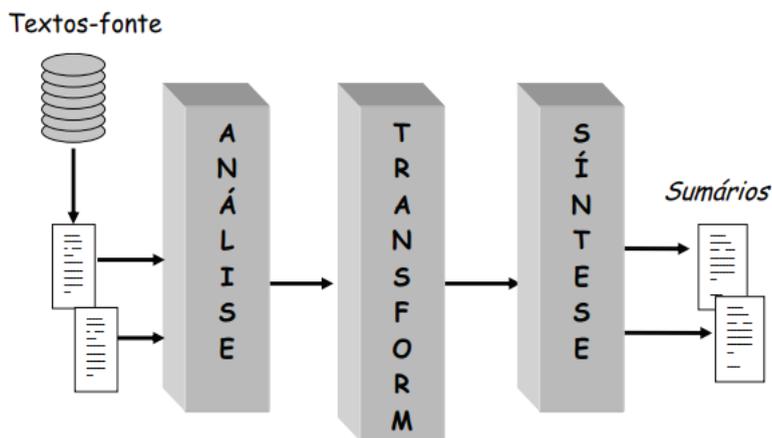
Diante das inúmeras produções textuais que povoam a web o processamento desses documentos é marcado pela irrelevância numa simples busca que demanda tempo de leitura e resposta para aquilo que se procura, a sumarização surge como um mecanismo de análise automática do conteúdo de um documento a fim de saber a ideia principal. Este processo remete a representação de um documento do qual se extraiu um resumo. Neste caso o resultado da sumarização remete a um texto-fonte, o que basicamente assemelha-se a extração de termos significativos que

representam conceitos contidos no documento. Deste modo Rino e Nunes (2005), ressaltam que sumários são textos produzidos a partir de textos, podendo servir, principalmente, de indexadores ou substitutos dos mesmos. Destacam ainda que os métodos de extração da sumarização automática consistem em:

- Identificar segmentos textuais relevantes para compor um sumário;
- Extrair do texto-fonte as unidades mínimas de significado que incluam tais segmentos;
- Justapor cada uma dessas unidades, resultando no sumário final.

A sumarização prevê minimamente uma ordem concisa de apresentação do texto fonte, representado por meio de elementos introdutórios, explicativos e conclusivos. Isto consiste basicamente a apresentar os pontos principais que tratam do conteúdo do texto-fonte a representação desta técnica pode ser ilustrada na Figura 2

Figura 2 – Arquitetura geral de um sumarizador automático



Fonte: Rino e Pardo (2010)

Existem duas abordagens principais para a sumarização automática a profunda e a superficial. Segundo Martins et al (2001), a profunda concentra-se no âmbito das teorias linguísticas e formais. E a superficial utiliza métodos estatísticos e experimentais. A Sumarização contenta-se em identificar as informações relevantes que irão compor o resultado final. Além disso, Balage Filho (2007, p.4), destaca que existem dois tipos básicos de sumários: o extrato e o *abstract*.

O extrato é um resumo produzido extraindo-se do texto-fonte frases que expressem as ideias principais, ao contrário do *abstract*, uma forma distinta de apresentar as mesmas ideias que o autor do texto desejava expor ao leitor. Os extratos são, em geral, produzidos por métodos da abordagem superficial. Os *abstracts*, por sua vez, podem ser produzidos na abordagem profunda.

Para Cabral (2015), as técnicas de sumarização extrativas produzem um subconjunto das sentenças mais importantes de um documento, exatamente como elas aparecem no documento original. E as abstrativas visando auxiliar e melhorar a coerência entre as sentenças, eliminando redundâncias e deixando clara a relação entre as frases.

Identificar as informações principais de documentos no ambiente web é uma tarefa que demanda tempo, por isso a sumarização surge como uma alternativa que diminui o tempo de análise de conteúdo de documentos. Segundo Pardo (2008, p.2):

O uso de sumários pode melhorar certos aspectos da recuperação de informação e da categorização de textos. Sumários também podem ajudar nas questões de usabilidade de interfaces e de acessibilidade e inclusão digital.

Nesta perspectiva, o uso de um sumariador para textos disponíveis na web consiste basicamente em poupar tempo de quem precisa ler grande quantidade de textos para se extrair a ideia principal do mesmo. Isto surge como uma alternativa principalmente para quem precisa tratar grandes massas documentais em determinado domínio. Beneficiando-se tanto de ideias principais para saber do que se trata o documento como da extração de descritores que possam representar os conceitos contidos nas fontes textuais.

Mapa Mental e Conceitual

O uso de representações sistêmicas para organização do conhecimento é expresso de diversas formas, seja por meio de diagramas, esquemas, mapas mentais, mapas conceituais e ontologias. Identificar o conteúdo e extrair representações conceituais que possuem relações entre si objetivam alavancar os processos de aprendizagem e apropriação do conhecimento. De acordo com Marques (2008, p.28).

Os mapas mentais (termo traduzido do inglês "*mind maps*") e os mapas conceituais ("*concept maps*", no original) são representações esquematizadas de informação que permitem demonstrar facilmente relações de significado e de hierarquia entre ideias, conceitos, factos ou acções.

Ressalta ainda que, estas “ferramentas de organização de informação” recorrem a uma espécie de simbiose entre linguagem gráfica e textual. Deste modo propõe uma forma de leitura diferente cujo objetivo é expandir as relações conceituais de um texto, tendo em vista que o conceito caracteriza-se como unidade básica do conhecimento. Estes arranjos proporcionam uma visão aprofundada do conteúdo dos documentos.

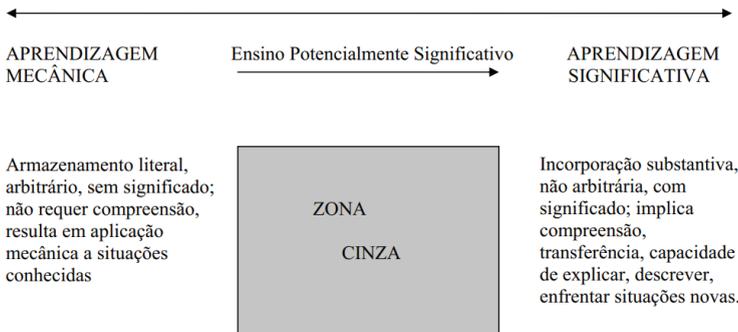
Um mapa mental propõe uma leitura não linear. Hermann e Bovo (2005), concebe o mapa mental essencialmente como um diagrama hierarquizado de informações, no qual é possível facilmente identificar as relações e os vínculos entre as informações. Estas perspectivas de leituras reinventam as formas de apreensão do conhecimento, deste modo Marques (2008, p.28), afirma que:

Fomos educados no sentido de pensar de forma sequencial, aprendendo a ler “de cima para baixo e da esquerda para a direita”. Fragmentamos os problemas e as questões com que nos defrontamos em pequenas partes. Essa metodologia, comprovadamente útil, impede-nos contudo de ter uma boa visão do “todo” e das possíveis relações entre todos os aspectos de uma determinada questão.

Para Oliveira (2006), os mapas mentais são meios de estruturar e armazenar conhecimento. Seu uso implica basicamente na organização e hierarquização do conhecimento do indivíduo, que de forma categórica estabelece relações entre si facilmente memorizável. Neste sentido dentro da teoria de aprendizagem, um eixo subjacente do cognitivismo que se destaca é a aprendizagem significativa, de David Ausubel que em suma explica que uma informação ao ser incorporada no estoque mental de um indivíduo e achando relações com informações existentes, gera uma relação conceitual a partir dos significados que estas informações estabelecem.

Na aprendizagem significativa é destacada a aprendizagem mecânica ou memorística que caracteriza-se pela pouca influência e falta de significação entre as relações conceituais, a Figura 3 apresenta uma visão da transição da aprendizagem mecânica para significativa.

Figura 3 – Visão da aprendizagem mecânica para significativa

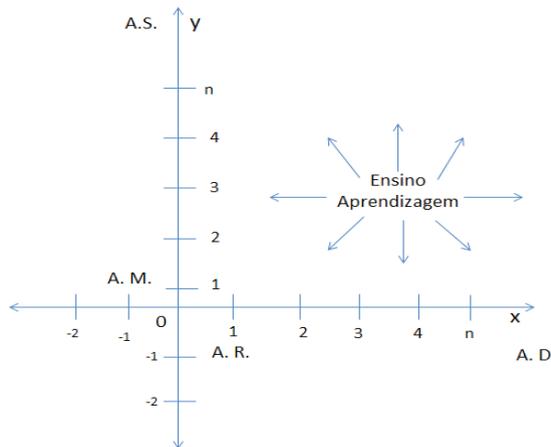


Fonte: Moreira (2012)

Moreira (2012), destaca que a Aprendizagem Significativa (AS) e Aprendizagem Mecânica (AM) não constituem uma dicotomia: estão ao longo de um mesmo contínuo e que, há uma “zona cinza” entre elas. Por isso a Aprendizagem Significativa (AS) torna-se algo progressivo, não sendo portanto natural ou automática.

Outros pontos destacados referem-se à Aprendizagem Receptiva (AR) cuja informação encontra-se pronta para incorporar-se a um conhecimento, não cabendo nenhum tipo de reflexão e a Aprendizagem por Descoberta (AD) que provoca um busca precípua daquilo que se pretende aprender. Na Figura 4 é possível perceber as relações desses tipos de aprendizagem.

Figura 4 – Relações entre AS, AM, AR e AD



Fonte: Adaptado de Moreira (2012)

Adotando uma comparação cartesiana e considerando os eixos x e y , quanto mais próximo de 0 (zero), maior a tendência para a aprendizagem mecânica e receptiva e quanto mais próximo de n maior a possibilidade de aprendizagem significativa e por descoberta. Isto implica dizer que um a aprendizagem positiva > 0 (maior que zero) proporciona n relações entre informações que se tornarão conceitos e conseqüentemente conhecimento. Nesta perspectiva o que se pretende destacar neste trabalho dentro da aprendizagem significativa é o que Moreira (2012) chama de aprendizagem representacional relacionada a um segundo tipo de aprendizagem significativa, a aprendizagem conceitual.

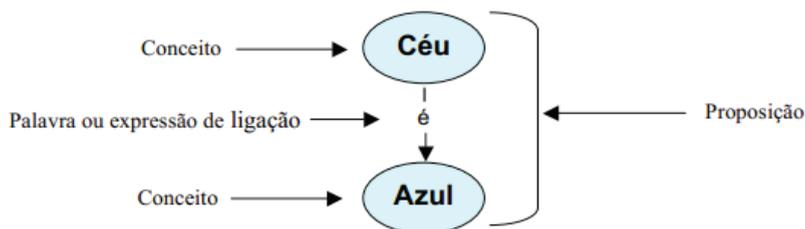
A aprendizagem conceitual ocorre quando o sujeito percebe regularidades em eventos ou objetos, passa a representá-los por determinado símbolo e não mais depende de um referente concreto do evento ou objeto para dar significado a esse símbolo. (MOREIRA, 2012, p.16).

A aprendizagem conceitual envolve a categorização de elementos com determinados padrões extraídos de objetos, a partir disto é necessário verificar os conceitos sobre estes e suas possíveis relações. A propósito dentro da definição incisiva de conceito orientada a unidade do conhecimento apresentada por Dahlberg (1978), Ausubel; Novak; Hanesian (1980) afirma que conceitos consistem na abstração dos atributos essenciais que são comuns a uma determinada categoria de objetos, eventos ou fenômenos. Esses atributos essenciais acabam sendo elementos representativos que possuem relações entre si.

Neste sentido os mapas conceituais surgem como uma apresentação gráfica dessas relações conceituais que segundo Machion (2007), foi originalmente desenvolvida em 1972 durante um projeto de pesquisa de Joseph Novak na Universidade de Cornell. Estas relações entre conceitos possui um elo intitulado por Ausubel; Novak; Hanesian (1980) de subçunção.

A partir de representações gráficas de conceitos com ligações entre si e apresentadas de forma hierárquica obtém-se um mapa conceitual. Na Figura 5, é apresentada uma forma básica para constituição de um mapa conceitual formado por duas palavras que caracterizam-se como substantivo e adjetivo entendidos como conceitos e uma palavra ou expressão de ligação que indicam uma afirmação.

Figura 5 – Forma básica de um mapa conceitual



Fonte: Marques (2008)

Neste ponto é necessário esclarecer que a construção de um mapa conceitual por vias humanas a partir de um texto envolve a abstração de conceitos extraídos e relacionados entre si. Com o estoque mental de conceitos predominantes no indivíduo para incorporação de um novo conhecimento. Dentro dessa lógica do conhecimento informal para o conhecimento formal entende-se a como a conversão da linguagem natural (informal) para a linguagem de máquina (formal).

Nesta seção foram apresentadas algumas definições sobre PLN e seu uso para extração de termos que podem ser analisados por meio de sua estrutura morfológica e léxica, mapas mentais como representações esquematizadas de conceitos a relação da aprendizagem significativa e sua relação com a construção de mapas conceituais. Isto teve como objetivo apresentar os métodos de representação da informação e organização do conhecimento de uma visão que apesar de fluir na tecnologia têm como modelo a complexidade mental de um indivíduo, além disso, a possibilidade de gerar um encadeamento para se entender tecnologias de inteligência que influenciam e modificam as formas de pensar.

INTERFACE WEB (SUMAPS)

Neste trabalho a interface web proposta tem por nome SUMAPS do acrônimo *Summarize for Maps* que traduzido significa “resuma para mapas”, pois esta é a ideia principal, gerar um resumo automático a partir de um texto disponível na web, extrair as palavras-chave e com estas representar uma espécie de mapa mental. Para isto recorreu-se a literatura científica sobre organização do conhecimento, teoria do conhecimento, análise de conteúdo, indexação, sumarização, mapas conceituais para nortear os eixos de sustentação desta concepção.

A geração automática de um resumo é extraída de um determinado texto, que foi construído a partir de uma lógica conceitual. Ele pode ser representado a partir de seus descritores. O SUMAPS não apresenta como

resultado o conceito estrito sobre determinado objeto descrito no texto, mas indica possibilidades que futuramente poderão ser aprimoradas.

Grande parte das páginas webs disponíveis hoje encontram-se no formato *Hyper Text Markup Language* – HTML, devido a sua fácil manipulação e edição do script que compõe a página, que pode incorporar formatos como o – *eXtensive Markup Language* – XML, *Java Script Object Notation* - JSON e funcionalidades embutidas com o PHP (*Hypertext Pre-processor*). Estes formatos permitem a criação de páginas e aplicações webs dinâmicas que exijam requisições e respostas desses formatos entre si para se chegar a um produto final, que pode ser desde um formulário simples, a um sistema de gerenciamento de conteúdo – CMS (*Content Mananger System*).

O SUMAPS foi escrito em PHP e incorporado à uma página HTML, a escolha do formato PHP se justifica pelo desempenho de um formato adequado a web. Leite (2014) concebe o PHP: como uma coleção de *supertags* de HTML, que permite adicionar funções do servidor às páginas da Web. Destaca ainda que grande parte do que o PHP realiza é invisível ao usuário final. E que ao visualizar uma página de PHP, não somos capazes de afirmar que não foi escrita em HTML, pois o resultado final do PHP é HTML.

O PHP é uma linguagem de *script Open Source*, este formato é extremamente útil, pois tem suporte a diversos tipos de bancos de dados, como dBase, Interbase, mSQL, MySQL, Oracle, Sybase, PostgreSQL entre outros. Além de ser possível construir uma página baseada em um banco de dados, possui suporte a outros serviços através de protocolos como HTTP, IMAP, SNMP, NNTP, POP3 entre outras funcionalidades.

O PHP é um projeto sob a proteção da Apache Software Foundation e possui código-fonte aberto. Leite (2014, p.8) destaca que:

O PHP é um módulo oficial do servidor HTTP Apache. Isso significa que o mecanismo de script do PHP pode ser desenvolvido no próprio servidor Web, tornando a manipulação de dados mais rápida. Assim como o servidor Apache, o PHP é compatível com várias plataformas, o que significa que ele executa em seu formato original em várias versões do UNIX e do Windows.

Devido a todos estes requisitos funcionais, a compatibilização com o ambiente web, e sistemas operacionais usuais, como o Windows e Linux que possuam um navegador web popular é que se optou basicamente por este formato. Apesar dos comandos em PHP estarem incorporados numa página no formato HTML, exige-se basicamente elementos básicos de programação em PHP para execução das funções atribuídas à página.

Figura 6 – Exemplo de uma página em HTML com PHP embutido

```

<!DOCTYPE HTML>
<html>
  <head>
    <title>Exemplo</title>
  </head>
  <body>
    <?php
      echo "script em PHP";
    ?>
  </body>
</html>

```

Fonte: Sousa (2017)

Para se fazer uma chamada em PHP numa página HTML basta inserir o comando “<?php” para iniciar e “?>” para encerrar, ao final de cada expressão deve-se inserir “;” (ponto e vírgula). Como se trata de uma linguagem que lida com expressões. Notadamente deve-se atribuir um valor a constantes e variáveis, ou seja, se escrever que “\$b=name” então todas as vezes que escrever \$b será como se estivesse escrevendo “name”. O HTML serve para estruturar a página e o PHP se ocupa da parte invisível que pode conter várias funções a serem executadas. Sabendo disto e de outros elementos de lógica de programação o SUMAPS se beneficia do PHP para fazer várias chamadas a *Applications Programing Interface* – APIs que executarão as funções aqui propostas.

As instruções de como fazer a chamada da API estão no próprio site da mesma que a disponibiliza, desde que a pessoa possua uma chave de acesso mediante cadastro, esta chave é inserida no script e gera um número limitado de acesso a API. Isso permite a sumarização, extração de palavras chave e visualização. Em suma isto quer dizer que basta colar o link com endereço no campo de busca na página do SUMAPS que ele vai executar tudo isto em alguns segundos.

Esta proposta apresenta um passo inicial para aplicação de tecnologias voltadas ao auxílio do processo de organização do conhecimento e indexação na biblioteconomia. Não é necessário portanto um conhecimento avançado em linguagem de programação, porém de elementos norteadores como as APIs e uma lógica que possa integrá-las. O processo

de criação e uso de tecnologias pelo bibliotecário deve ser algo que objetive a melhoria do desenvolvimento de suas atividades, pois ninguém conhece mais a profissão quanto os que a exercem ativamente, por isso devemos estar à frente de criações que interfiram diretamente em nossa prática profissional.

O uso de APIs em Softwares (SMMRY)

Programas executados no ambiente web operam um pouco diferente de programas *off-line*, que normalmente são instalados em um computador, entretanto ambos podem se beneficiar de instruções pré-ordenadas, que dispensarão o uso de outros *softwares*, pois isto possibilita a ampliação de funções que podem ser executadas simultaneamente. O uso de APIs é bastante comum em aplicações web. Segundo Pinotte; Cury; Zouaq (2015, p.175):

Grandes companhias do ramo de tecnologia da informação (por exemplo, Facebook, Apple, Google, Twitter etc.) disponibilizam suas APIs (*Application Programming Interfaces*) de acesso aos seus serviços utilizando esse modelo de computação.

Dentre elas a API mais popularizada certamente é a do *Google Maps* que oferece um serviço de localização por meio de mapas que pode ser embutido dentro de uma página HTML de um site. Empresas como *Google* se beneficiam disso para ampliar diversas soluções para o ambiente web, como o *Google Ajax Search API*, que se trata de uma interface de programação que permite que você coloque uma caixa de pesquisa do Google em seu site e pesquisa personalizada em suas páginas. As APIs oferecem funcionalidades que permitem o uso de mais de uma na mesma página em HTML como é o caso do SUMAPS.

Neste caso, a API que executa a maior parte da função é a do sumariador, intitulado SMMRY que foi criado principalmente para reduzir textos elencando as frases mais importantes. Esta API executa suas tarefas por fases, por meio de um algoritmo que classifica as frases por importância, reorganiza o resumo e concentra um tópico por meio de palavras-chave, remove as frases de transição, as cláusulas desnecessárias e os exemplos excessivos. As etapas consistem em:

- a) Associar as palavras com os seus homólogos gramaticais.
- b) Calcular a ocorrência de cada palavra no texto.
- c) Atribuir cada palavra com pontos dependendo de sua popularidade.
- d) Detectar quais períodos representam o fim de uma frase.
- e) Dividir o texto em frases individuais.

- f) Gerar um ranking de frases pela soma de pontos de suas palavras.
- g) Retornar “X” frases mais altamente classificados em ordem cronológica.

Segundo Cabral (2015), o SMMRY foi criado em 2009 por Amir Elmaani, ressalta que esta ferramenta foi desenvolvida em PHP e funciona online tendo como entrada, arquivos de texto simples (*txt*) ou hipertexto (HTML), produzindo uma saída do mesmo tipo de arquivo. O SMMRY é acessível por meio de uma API e pode ser usado em aplicações que podem exigir um resumo de uma página web ou arquivos de textos variáveis, o número máximo de frases devolvidas chega a 40.

Para usar esta API basta inserir o resumo do script com a chave de acesso, a *API key* para páginas externas como é o caso do SUMAPS. Além disso, pode se inserir requisições opcionais, como o número de sentenças, palavras-chave, não inclusão de citações entre outras. As respostas dessas requisições são retornam um *JavaScript Object Notation (JSON)*, este formato é a chave dessa interface, pois ele que possibilita a visualização destas palavras chave no D3, merecendo portando uma explicação prévia do que se trata.

O JSON é um formato de intercâmbio de dados leve. É fácil para os seres humanos lerem e escreverem. É fácil para máquinas analisarem e gerarem, um formato de texto que é completamente independente do idioma, ele é construído em duas estruturas, uma coleção de pares nome / valor e uma lista ordenada de valores. Além disso, assume duas formas, de objeto e matriz.

- Um objeto começa com {(chave esquerda) e termina com} (chave direita), cada nome é seguido por: (dois pontos) e o nome / valor pares são separados por, (vírgula).
- Uma matriz começa com [(colchete esquerdo) e termina com] (colchete direito), os valores são separados por, (vírgula).

O JSON pode ser entendido como uma extensão do XML, sua estrutura é compreendida tanto por máquinas quanto humanos. A API do SMMRY retorna um JSON. Por meio do comando “*Json encode*” o D3 recebe e lê este arquivo. Isto é extremamente importante para o SUMAPS por que o SMMRY gera um JSON e D3 lê esse JSON, ambos são compatíveis com o PHP e operam no ambiente web.

Data – Driven Documents - D3

Uma das questões levantadas para a concepção deste trabalho se deve ao fato de visualizar textos em formatos dinâmicos e automáticos.

Isto exigiu uma busca por diversos softwares que pudessem fazer isto sem muita interferência humana. A resposta para essa busca veio pelo uso de outra API, a do *Data – Driven Documents* – D3 uma abordagem de representação transparente para visualização na web.

Com D3 é possível aplicar transformações dinâmicas, tanto gerar e modificar o conteúdo, por que ele se baseia em documentos orientados a dados. Isto garante visualizações nos mais diferentes modelos como, calendários, diagramas, mapas, pacotes hierárquicos, matriz de dispersão, bares agrupados e empilhados, clusters gráfico, entre outros (D3js,2016). Seus criadores são Michael Bostock, Vadim Ogjevetsky e Jeffrey Heer da Universidade de Washigton nos Estados Unidos

O D3 permite transformar documentos com base em dados. Além disso, altera um documento existente em resposta à interação do usuário, animação ao longo do tempo, ou notificação mesmo assíncrona a partir de um terceiro. Possui uma abordagem híbrida onde é possível que um documento inicialmente processado no servidor, e atualizado no cliente através de D3 não sobrecarregue nenhum dos dois. Com sobrecarga mínima, o D3 é extremamente rápido apoiando grandes conjuntos de dados e comportamentos dinâmicos.

Seu uso suporta navegadores “modernos”, exceto *Internet Explorer 8* e versões mais antigas. É compatível com *Firefox, Chrome, Safari, Opera, IE9+, Android* e *iOS*. O D3 também está disponível por meio de gerenciadores de pacotes, incluindo: *NPM (Node.js)*, e *Browserify, Component, Jam, Composer / Packagist (PHP), SPM, JSPM, NuGet (.Net), and AMD*. O D3 permite uma visualização dinâmica que por meio de arquivos no formato JSON ou CSV podem ser manipulados e visualizados, no SUMAPS o script em PHP para acionar a API do D3.

O SMMRY gera um JSON que é adaptado à estrutura do JSON do D3 para poder ser interpretado, isto é importante, pois cada modelo de visualização necessita de arquivos em formatos específicos como JavaScript, JSON ou CSV para carregar os dados a serem visualizados. No SUMAPS as palavras-chave são extraídas para comporem o JSON que será lido pelo D3, nisto será possível uma visualização dinâmica por meio da extração automática de palavras que constituirão um mapa.

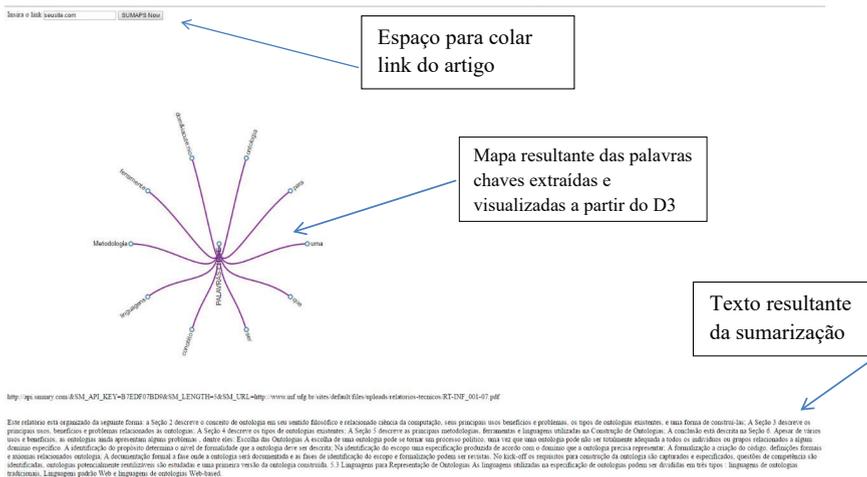
A compatibilização do D3 com a API do SMMRY foi o que permitiu a criação e interação dessa interface, pois as maiorias dos softwares disponíveis de criação de mapas mentais necessitam de grande interferência humana para compor um mapa, com isto é possível resumir os textos e gerar mapas simultaneamente de vários textos, com a interferência humana reduzida consideravelmente.

Uso e aplicação do SUMAPS

Para avaliação do uso do SUMAPS optou-se por um universo de 30 artigos científicos sendo que 20 são em língua portuguesa e 10 em língua inglesa, todos no formato PDF (*Portable Document Format*), os artigos em português foram extraídos da base de dados científica *Scielo* optando por textos recuperados com a palavra-chave ciência da informação. Dos artigos em Inglês, 5 foram extraídos da base de dados *Elsevier* recuperados pela palavra-chave “*ontologies*” e 5 extraídos de forma geral no *Google Scholar* com a mesma palavra chave.

A análise desses artigos reservou-se no tocante a identificar a quantidade de conceitos assim como o nível de precisão de sentido dos resumos gerados para atentar se existe de alguma forma a possibilidade de dispensar a interferência humana. Usando um método de avaliação automática por meio de lemas e variação lexical dos textos sumarizados. A API que sumariza o texto retorna basicamente três elementos, o título contido no URL, o conteúdo e as palavras-chave que aparecem mais vezes. A lista de artigos e seus respectivos resumos encontram em anexo para maior detalhamento. A aplicação do SUMAPS deu-se por meio da colagem dos *links* e compilação dos textos sumarizados, conforme a Figura 7.

Figura 7 – SUMAPS Modelo



Fonte: Sousa (2017)

A API do SMMRY opera em 7 etapas conforme exposto na seção 4.1 e basicamente consistem em identificar afirmações, definições e conceitos. As explicações e descrições são atreladas a ligamentos do texto que nor-

malmente apareceram no meio ou no final do resumo, por meio de frases. Observa-se que a API não isola apenas termos, mas sentenças que direcionem a persistência da ênfase do autor nas frases. Na figura 20 verifica-se que o resumo construído pelos autores do trabalho informa apenas o objetivo e uma palavra chave.

O exemplo mostra basicamente como o SUMAPS funciona, com o uso da API do SMMRY nota-se a que a leitura total do documento possibilitou a identificação das seções que compõe o documento e associá-las a ordem lógica do texto. As palavras-chave não possuindo um filtro para corrigir questões de acentos, preposições e artigos aparecem com caracteres distorcidos ainda.

A análise preliminar consistiu em identificar primeiramente do que se tratava o documento por meio do assunto tendo em vista que nem sempre o título prescreve o conteúdo do texto, posteriormente verificou-se o nível de precisão do resumo para detectar sua acurácia. O tipo de resumo gerado caracteriza-se como informativo, pois destaca alguns pontos principais do texto-fonte e necessita, portanto da consulta do mesmo para obter maiores detalhes.

Os textos com maior precisão sobre assunto possuíam uma abordagem explicativa, sendo que em português o número mínimo de páginas sumarizada por texto foram 3 e o máximo 22 de páginas em inglês o mínimo foram 9 e o máximo 54 páginas. Por padrão no site da própria API optou-se por retornar apenas 7 frases em cada texto, apesar de ser possível modificar o número de frases e 10 palavras-chaves. As frases contêm normalmente afirmações, explicações, descrições, definições ou conceitos.

Neste trabalho optou-se por identificar a porcentagem da variação lexical nas frases, pois se entende que estes são decorrentes de uma concentração de palavras em determinado domínio e sua variação diminui redundâncias e conseqüentemente os textos submetidos à sumarização podem conter unidades elementares para organização do conhecimento, fundamentado na própria visão de Bardin (1977), que afirma sobre a possibilidade de uma análise dos significantes (análise léxica, análise dos procedimentos), como forma de se analisar o conteúdo do texto analisado. A princípio, o SUMAPS precisa ser analisado por pesquisadores e profissionais que se interessem por este tema a fim de aperfeiçoá-lo.

Análise dos textos sumarizados e informados pelos autores

A análise dos textos sumarizados exige métodos e ferramentas que possam avaliar esse nível de automaticidade, por isso a dificuldade em encontrar métodos manuais e automáticos, permitiu apenas uma análise preliminar sobre o léxico textual especificamente no campo lexical que se aplica a determinado domínio.

O léxico constitui um conjunto aberto em qualquer sistema lingüístico e, por conseguinte, não apenas vastíssimo quando comparado com outros setores e níveis da língua (fonológico, morfológico, sintático) mas também em permanente expansão sobretudo numa língua de civilização. Por essa razão, o quantitativo é uma das propriedades fundamentais do vocabulário. (BIDERMAN, 1984, p.17).

Entende-se que a análise lexical é a mais viável para este trabalho, pois isto possibilita a relação destas palavras em determinado campo do conhecimento. A análise semântica que trataria do significado textual depende de um estudo mais aprofundado, todavia é possível fazer esta análise com softwares como o *Recall-Oriented Understudy of Gisting Evaluation* – ROUGE, um pacote de software para avaliação automática de resumos, desenvolvido por Chin-Yew Lin, enquanto estava no Instituto de Ciências de Informação da Universidade do Sul da Califórnia (USC / ISI) e o projeto *Freeling* criado por Luís Padró no Grupo de Processamento de Pesquisa de Linguagem Natural UPC da Universidade Politécnica da Catalunha.

Para análise destes textos sumarizados foi necessário o uso de uma ferramenta *online* chamada Linguakit da *Cilenis Language Technology*, empresa que atua no campo do Processamento da Linguagem Natural (PLN) e que tem algumas de suas funções baseadas no Projeto *Freeling*. Este, mecanismo apresenta as seguintes funções:

- a) Análise completa das funções oferecidas
- b) Resumidor de textos
- c) Conjugador verbal
- d) Frequência de palavras
- e) Palavras-chave em contexto
- f) Etiquetador morfossintático
- g) Analisador sintático
- h) Analisador de sentimento para saber se um texto exprime opiniões positivas ou negativas
- i) Identificador de idioma
- j) Extrator de palavras-chave

- k) Extrator multipalavra, ou seja palavras que precisam de outras para conter significação
- l) Reconhecedor de entidades que extrai de um texto as entidades, nomes próprios e palavras ou dados numéricos.
- m) Supercorretor
- n) O extrator de tripletas semânticas: sujeito-relação-objeto

A função optada foi à análise completa por meio de dados estatísticos a respeito dos textos sumarizados, a figura 21 apresenta um exemplo de resumo estatístico após o processamento. Onde é apresentado o número de frases no resumo, a variedade lexical, que aponta se o texto é redundante ou não, número de palavras, de caracteres, lemas¹ diferentes que significam ideias ou enunciados contidos no texto. E por fim caracteres sem espaços.

O Linguakit trabalha com atribuições de valores a todas as palavras encontradas nos textos, em outros softwares isso corresponderiam aos unigramas, bigramas, trigramas e n-gramas das palavras analisadas. Entretanto sua interface amigável atribui ao que chama de peso a cada palavra e a soma desses pesos geram as classificações. Apesar de o Linguakit gerar resumos ele não lê arquivos no formato PDF nos planos grátis e de usuários cadastrados. Diferente do SUMAPS que faz a leitura e a sumarização de arquivos no formato PDF disponíveis na web. Por isto o uso do Linguakit não substitui o SUMAPS por que este último não tem funções avaliativas de textos, realizando portanto a sumarização dos artigos e os textos resultantes submetidos à análise com o Linguakit, posteriormente foram extraídos os resumos contidos nos artigos selecionados que notadamente foram elaborados pelos autores dos mesmos e também submetidos ao mesmo tipo de análise.

Em síntese, dos 20 textos sumarizados em língua portuguesa foi possível identificar uma variação lexical média entorno de 73,55 %, o número de frases manteve-se entre 5 e 7 e os lemas diferentes ou palavras variou entre 85 e 204 palavras que representam sentenças ou enunciados, isto pode ser melhor descrito no quadro 3.

Na análise dos resumos informados pelos autores dos artigos, nota-se uma variação entre 1 a 8 frases, lemas entre 11 a 119 e uma variação lexical de 60,61% a 100%. Os textos 3, 5 e 12 não puderam ser analisados, pois geravam erros e não apareciam os resultados mesmo com inúmeras tentativas.

¹ Preceito escrito; sentença, palavra que figura como entrada num dicionário ou num vocabulário. **Dicionário da Língua Portuguesa com Acordo Ortográfico**. Porto: Porto Editora, 2003. Disponível em: < <http://www.infopedia.pt/dicionarios/lingua-portuguesa/lema> >. 29 jul 2016.

Percebeu-se que o nível de variação léxica indicava, que há pouca redundância no texto sumarizado. Dos textos resultantes da sumarização, 9 apresentaram uma taxa de 70% de variação lexical, 4 estavam na casa dos 80% e 6 com 60% sendo que o T1² tinha uma variação léxica de 60,61 % no resumo informado pelo autor e ficou com 68,31% no SUMAPS, o que equilibra a comparação, mas ambos abaixo de 70%. Desse total a taxa de equilíbrio da variação lexical fica entorno de 70% contando com T1, pois foi contabilizado o nível de variação em relação aos resumos dos autores. O que indica um nível aceitável nesta pequena amostra desse universo textual.

Quanto aos textos em língua inglesa os 5 primeiros extraídos da base de dados *Elsevier* apresentaram um alto nível de precisão devido ao fato do sumariador incorporar o resumo dos artigos elaborados pelo autor ao resultado da sumarização, apesar disso com exceção de T5 a variação lexical no SUMAPS foi superior aos resumos informados pelos autores. Os espaços deve-se a falta de resumos informados pelos autores nos textos sumarizados. Por isso não é possível explicitar um quadro comparativo entre todos os textos.

Nota-se ao todo que no SUMAPS apesar de apresentar uma truca-gem inicial por apresentar conceitos e sentenças isoladas é possível identificar do que se trata o resumo, só não há uma concatenação exata das ideias apresentadas. Dentre as dificuldades percebe-se a relação que o sumariador faz com figuras e gráficos contidos nos textos sumarizados, porém isto não é possível de visualizar no resumo. Quanto à questão do idioma notadamente o inglês é mais aceito devido à exigência do próprio idioma de não existir dois elementos de negação em uma mesma frase, por exemplo, diferente do português.

Apesar disto o nível de aproveitamento do resumo ainda é bom por conter na maioria das vezes conceitos e afirmações explicitadas nos textos, além disso, a taxa de compressão dos textos é muito alta, pois no total o número de páginas submetidas a sumarização variaram de 3 a 54. Quanto às palavras chave extraídas em português foi constante a presença de preposições e artigos diferente dos textos em inglês que retornaram palavras expressivamente significativas.

Os mapas mentais, assim intitulados são palavras extraídas que são graficamente apresentadas com conectivos, visualizando desta forma a representação do texto sumarizado, em português o aproveitamento foi menor em relação ao inglês, as apresentações gráficas dão suporte às representações textuais expresso nas frases. Consequentemente é possível

² T seguido de um número é uma notação atribuída em ordem aos textos sumarizados em português e inglês respectivamente.

identificar elementos que podem ser utilizados para uma possível busca de termos extraídos do próprio texto. O resumo e as representações por palavras-chave, o mapa mental direcionam para a compreensão do assunto do texto sumarizado por meio de seu conteúdo.

O nível de análise da sumarização do SUMAPS caracteriza-se como superficial pois utiliza ainda métodos estatísticos e experimentais, sendo portanto extratos que recuperam conceitos e sentenças dos artigos, apesar disso apresenta características de abstracts que possuem uma abordagem mais profunda, pois em alguns casos, ele constrói um novo arranjo frasal de acordo com a popularidade das sentenças.

CONSIDERAÇÕES FINAIS

A linguagem é um recurso por meio do qual se transmite esse conhecimento por meio de informações. Os meios de representação do conhecimento podem facilitar sua apreensão ou não. O uso de tecnologias digitais nos últimos anos tem facilitado a comunicação entre as pessoas e as máquinas, por meio das quais desenvolvemos simulacros da vida humana através da inteligência artificial, que visa mais do que isso, como a própria aprendizagem artificial. Neste meio, o Processamento da Linguagem Natural avança para o aprimoramento do domínio de elementos de representação simbólica da linguagem natural para artificial.

Neste sentido a tentativa de entender o que contém a linguagem escrita, por meio de textos com informações que representam um dado conhecimento, indica os passos a serem tomados para potencializar o aprendizado de máquina. Onde a redução de ruídos facilitará primeiramente, a recuperação da informação desejada, além disso, possibilitarão a expansão de novas pesquisas neste ramo.

A sumarização de textos apresenta-se nessa perspectiva como ferramenta para geração automática de resumos significativos contendo conceitos, afirmações, explicações e descrições do conteúdo dos textos submetidos para tal fim, indica também a necessidade de aprimorar esse tipo de ferramenta para usos acadêmicos e pessoais com maior eficiência, atrelada a outras ferramentas que potencializem isso.

A forma de apresentação do que é submetido a esse tipo de processamento destaca o modo como o usuário vai lidar com as informações resultantes que expressam um determinado conhecimento, sentenças longas e truncadas deixam a desejar pelo próprio nível de leitura, apresentando as palavras-chave como alternativa de representação por meio de possíveis termos. A visualização, por meio de elementos conectivos,

indica as relações que estas palavras possuem no texto e dentro de um domínio.

Textos originariamente digitais devem ser os primeiros a serem analisados uma vez que, as ferramentas digitais na maioria das vezes são pensadas para processar elementos digitais, mesmo que estes representem elementos materiais. Pelo vasto número de textos digitais o beneficiamento da análise por meio deles deve-se a facilidade de encontrá-los, sendo adequado ou não a quem procura.

A otimização na busca e eficiência da recuperação de informações que representam conhecimentos no mundo digital especificamente da web exigem um elevado investimento em tecnologias que permitam isso. O domínio das técnicas e possibilidades de aperfeiçoamento, levam na maioria das vezes, iniciativas que liberam o domínio das técnicas de uso dessas tecnologias para todos que se interessarem, como é o caso das APIs.

O domínio de técnicas de formatação que possibilite formular e apresentar um produto tecnológico não é de domínio de todos os profissionais de todas as áreas, entretanto este domínio pode melhorar significativamente o desenvolvimento da área de atuação em questão. Pois isto gera o diferencial e valorização daquilo que se faz, seja por meios artificiais ou manuais.

Os entraves de domínio tecnológico, neste caso a programação em linguagem artificial, revelam os desafios dos profissionais que ingressam nessa área, porém isso não quer dizer que é impossível ou futurístico. Quer dizer que os olhos humanos estão voltados sim para o futuro, realmente, pois do passado já viemos a ação presente está mediando isto em cada avanço ou retrocesso.

Exigir de uma máquina, a apresentação semântica, respostas dignas de um ser humano com alta instrução é muito presunçoso atualmente, apesar dos avanços significativos, inclusive com robôs virtuais de conversação, entretanto devem ser considerados os erros para que se possa melhorar o desempenho.

Deste modo os resultados apresentados no processo de sumarização indicam um bom nível de aproveitamento dos textos em português obtendo um desempenho aceitável, nos textos em inglês, apesar de 40% não se qualificar para análise comparativa, o resultado final de desempenho mostra um nível de entendimento para identificação e concepção dos resumos.

As palavras-chave e o gráfico apresentado como mapa precisam de aperfeiçoamento para destacar a representação do conhecimento e estimulação da apreensão do mesmo, cabendo ressaltar que, quanto menor for a quantidade de palavras, mas simples será o mapa, e quanto maior for o número de palavras mais complexo e semelhante será em relação ao modelo original.

Em relação à construção dos resumos informados pelos autores, os periódicos científicos obedecem à orientação da ABNT na NBR 6028 (2003), que indica que este tipo de resumo deve conter de 100 a 250 palavras. Porém nos próprios resumos dos autores em língua portuguesa 5 tinham no máximo 100 palavras, caracterizando apenas como resumos breves e não científicos, apesar de estarem indexados na base do Scielo. Os resumos do SUMAPS variam de 185 a 433 palavras devido uma oscilação no número de frases, mas nenhum inferior a 100 palavras.

Esse desempenho apesar de satisfatório foi analisado apenas por um tipo de sumariador, sendo que há outros tipos. No entanto voltando ao objetivo deste trabalho, optou-se pela viabilidade e compatibilidade do mesmo tendo em vista que os outros pesquisados, não atenderam estes requisitos para concepção da interface por meio de APIs.

Portanto conclui-se que iniciativas como estas devem ser tomadas necessariamente por profissionais da informação, pois estes conhecem suas necessidades melhor do que ninguém e que o aporte teórico incorporado na formação deve ser um elemento basilar na concepção de artefatos para sua área de atuação e combinação com aquilo que apraz este profissional, seja a área tecnológica, social, cultural entre outras refletirá os avanços sofridos em cada área. A sociedade tem exigido respostas de todas as instituições mantidas por ela, assim como de quem as compõem. Este trabalho é uma das respostas que devem ser apresentadas não só a comunidade científica.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12.676**: Métodos para análise de documentos – determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro, 1992.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6028**: Informação e documentação - Resumo - Apresentação. Rio de Janeiro, 2003.

AUSUBEL, D.; NOVAK, J. D.; HANESIAN, H. **Psicologia educacional**. 2. ed. Rio de Janeiro: Interamericana, 1980. 625 p.

ÁLVAREZ, A. C. **Extração de informação de artigos científicos**: uma abordagem baseada em indução de regras de etiquetagem. Dissertação de mestrado, São Carlos: ICMC-USP, 2007.

BALAGE FILHO, P. P. et al. **Sumarização Automática de Textos Científicos**: Estudo de Caso com o Sistema GistSumm. NILC-TR-07-11, 2007. Série de Relatórios do Núcleo Interinstitucional de Linguística Computacional NILC - ICMC-USP.

BARDIN, L. **Análise de conteúdo**. Lisboa, Edições 70, 1977.

BIDERMAN, Maria Tereza Camargo. **A ciência da Lexicografia**. Alfa, São Paulo, 1984. Disponível em: < <http://seer.fclar.unesp.br/alfa/article/download/3676/3442> > Acesso em 29 jul 2016.

BIREME. **Manual de Indexação de Documentos para a Base de Dados LILACS**. BIREME. São Paulo: BIREME / OPAS / OMS, 2008. 392 p.

CABRAL, L. de S. **Uma plataforma para sumarização automática de textos independente de idioma**. 2015. Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Pernambuco, Recife 2015.

CENDÓN, B. V. Bases de dados de informação para negócios. **Ci. Inf.**, Brasília, v. 31, n. 2, p. 30-43, maio/ago. 2002.

DAHLBERG, I. Teoria do conceito. **Ciência da Informação**, Rio de Janeiro, v.7, n.2, 1978. p.101- 107.

GOMES, H. E. O indexador face às novas tecnologias de informação. **Transinformação** v.1 n.2, maio/ago,1989. p. 161-171.

GUINCHAT, C.; MENOU, M. **Introdução geral às ciências e técnicas da informação e documentação**. Brasília: FBB e IBICT, 1994. 540 p.

HERMANN, WALTHER; BOVO, VIVIANI. **Mapas mentais enriquecendo Inteligências**: Manual de Aprendizagem e Desenvolvimento de Inteligências: captação, seleção, organização, síntese, criação e gerenciamento de conhecimentos, 2005.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2ª ed. Brasília: Briquet de Lemos. 2004. 452 p.

LEITÃO, D. A. **NLForSpec**: uma ferramenta para geração de especificações formais a partir de casos de teste em linguagem natural. Recife, 2006. Dissertação (mestrado) - Universidade Federal de Pernambuco, 2006.

LEITE, A. C. C. R. **Programador Web**. Professionet Cursos Ltda. Rio de Janeiro: Professionet, 2014.

LEIVA, Isidoro Gil. Aspectos conceituais da Indexação. In: LEIVA, I. G.; FUJITA, M. S. L. **Política de indexação**. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012. 260p.

LIMA, G.A.B. Interfaces entre ciência da informação e ciência cognitiva. **Ci. Inf.**, v.32, n.1, 2003, p. 77-87.

MACHION, A. C. G. **Uso de ontologias e mapas conceituais na descoberta e análise de objetos de aprendizagem: um estudo de caso eletroestática**. Tese (doutorado em Matemática e Estatística) -Universidade de São Paulo, 2007.

MARQUES, A. M. de M. **Utilização pedagógica de mapas mentais e de mapas conceituais**. Dissertação apresentada à Universidade Aberta, Universidade Aberta, 2008.

MARTINS, C. et al. **Introdução à sumarização automática**. São Carlos:[s.n], 2001. Disponível em: < <https://www.icmc.usp.br/~tasparado/RTDC00201-CMartinsEtAl.pdf>>. Acesso em 12 jul 2016.

MENDES, M. T. P.; SIMÕES, M. da G. **Indexação por assuntos: princípios gerais e normas**. Lisboa: Gabinete de Estudos A&B, 2002.

MOREIRA, M. A. O que é afinal aprendizagem significativa? Aula Inaugural do Programa de Pós-Graduação em Ensino de Ciências Naturais, Instituto de Física, Universidade Federal do Mato Grosso, Cuiabá, 2010, **Qurriculum**, La Laguna, Espanha, 2012.

OLIVEIRA, N. A. da S. A educação ambiental e a percepção fenomenológica, através de mapas mentais. **Rev. eletrônica Mestr. Educ. Ambient.** vol. 16, jan./ jun. 2006. Disponível em: < <https://www.seer.furg.br/remea/article/view/2779> >. Acesso em 12 jul 2016.

PARDO, T. A. S. **Sumarização Automática: Principais Conceitos e Sistemas para o Português Brasileiro**. NILC - ICMC-USP: NILC-TR-08-04, Maio, 2008. Série de Relatórios do Núcleo Interinstitucional de Linguística Computacional.

PINOTTE, G. N; CURY, D; ZOUAQ, A. OntoMap: De Mapas Conceituais a Ontologias OWL. **Nuevas Ideas en Informática Educativa TISE**, 2015.

RINO, L. H. M.; PARDO, T. A. **A Sumarização Automática de Textos: Principais Características e Metodologias**. São Carlos: NILC/Departamento de Computação - UFSCar, 2010.

RINO, Lucia H. Machado; NUNES, Maria das Graças V. **Sobre Geração e Sumarização de Textos**. São Carlos: NILC - ICMC-USP, 2005.

Disponível em: < http://www.nilc.icmc.usp.br/nilc/download/NotasDidaticasICMC_67.pdf> Acesso em: 12 jun 2016.

RUBI, M. P. Os princípios da política de indexação na análise de assunto para catalogação: especificidade, exaustividade, revocação e precisão na perspectiva dos catalogadores e usuários. In: FUJITA, M. S. L. A; BOCCATO, V. R. C.; RUBI, M. P.; GONÇALVES, M. C. (Org.) **Indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias. Um estudo de observação do contexto sociocognitivo com protocolos verbais.** - São Paulo: Cultura Acadêmica, 2009.

SANTOS, G. C. **Fontes de indexação para periódicos científicos: um guia para bibliotecários e editores.** Campinas, SP: E-Color, 2011.

SILVEIRA, S. A. da. **Exclusão digital: a miséria na era da informação.** São Paulo: Fundação Perseu Abramo, 1 ed. 2 reimp., 2005.(Coleção Brasil Urgente).

SOUSA, Janailton Lopes. **Organização do conhecimento: interface web de sumarização e criação automática de mapas mentais com uso de APIs.** Monografia (Graduação), São Luis: UFMA, 2017.106 f.

TELINE, M. F. **Avaliação de Métodos de Extração Automática de Terminologia para textos em Português.** Dissertação (mestrado em Ciências Matemáticas e de Computação) - ICMC-USP, São Carlos, 2004.

VIEIRA, R.; LOPES, L. Processamento de linguagem natural e o Tratamento computacional de linguagens científicas. In: PERNA, C. L.; DELGADO, H. K.; FINATTO, M. J. **Linguagens especializadas em corpora: modos de dizer e interfaces de pesquisa.** Porto Alegre: EDIPUCRS, 2010. 255 p.

ZAVAGLIA, C. et al. **Avaliação de Métodos de Extração Automática de Termos para a Construção de Ontologias.** São Paulo: Núcleo Interinstitucional de Linguística Computacional NILC - ICMC-USP, 2005. Disponível em: <<http://www.nilc.icmc.usp.br/nilc/download/NILC-TR-05-01.pdf>> Acesso em: 07 jun. 2016.



A FOLKSONOMIA COMO TÉCNICA DE REPRESENTAÇÃO DA INFORMAÇÃO: estudo das práticas dos usuários da seção banco de cultura do Overmundo

Livia Regina Passos Silva

INTRODUÇÃO

Temos hoje um cenário da configuração informacional online que permite ao usuário da informação ser ao mesmo tempo o seu autor, editor e indexador, tendo o poder de influir no que é apresentado como o mais importante, tanto para ele como para outros usuários com interesses comuns. “A disseminação dessa forma colaborativa de classificação de informações é tal que alguns autores já sugerem uma nova revolução dentro da própria revolução que a internet representa” (LACERDA, VALENTE, 2007, p. 2)

Significantes como autoridade, instituição e usuário dentro da área de Biblioteconomia ficaram abstratos no contexto *web* de produção e uso da informação. Observamos nesse prisma que temos então, enquanto pesquisadores desta área que olhar com mais preponderância e atenção aos sujeitos que agem na *web*, interferindo e produzindo os conteúdos. Somente compreendendo suas ações é que, enfim, podemos entender as configurações atuais da dinâmica de construção, organização e uso do conteúdo *web*.

Dessa forma, percebemos a importância de estudar as práticas da Folksonomia utilizada pelos usuários colaboradores da seção banco de produtos culturais digitais do Site Overmundo.

Escolhemos trabalhar com o Site Overmundo em razão de seu caráter colaborativo e social, além de fazer uso da Folksonomia, quando o

usuário colaborador pode associar qualquer palavra a determinado registro e depois recuperar a informação através das *tags* (etiqueta) que ele mesmo criou.

Diante da exposição problematizada, justifica-se a escolha do tema não somente pela originalidade e atualidade, e/ou pelo interesse pessoal, mas também pela possibilidade de fazer um estudo de usuário gerado pelas seguintes questões:

- Como os colaboradores do Site Overmundo estão fazendo uso da Folksonomia para organizar seus conteúdos disponibilizados na net?
- Qual a relação da *tags* (etiquetas) com o conteúdo representado?
- Qual a motivação que sustenta o processo de etiquetagem do usuário?

Para tanto, delimitamos como objetivos analisar os critérios de indexação/etiquetagem usados pelos usuários colaboradores da seção banco de produtos culturais digitais do Website Overmundo para organizar seus conteúdos, assim como:

- Conhecer a relação das *tags* (etiquetas) com o conteúdo representado pelo usuário colaborador;
- Verificar a motivação que sustenta o processo de etiquetagem do usuário colaborador;
- Descrever o site do Overmundo;
- Classificar as etiquetas atribuídas a 6 (seis) produtos de 3(três) usuários da seção banco de cultura do Overmundo por meio da metodologia de análise de assunto para a indexação dos textos literários criada por Pejtersen (1978).

Diante desse contexto, discute o papel da representação do conhecimento e os esquemas de representação como instrumento desse processo por Brandt e Medeiros (2010), Carlan (2010), Piedade (1983), Dodebei (2002), Luger (2004), Oliveira e Carvalho (2008), Feitosa (2006), etc.; Introduce o tema Folksonomia, explanando sobre seus conceitos, tipologia, estrutura, críticas, utilização da etiquetas/*tags* e etiquetagem por Aquino (2007), Primo (2006), Wal (2007), Haykin e Moher (2008), Silva e Blattmann (2007), Brosch (2010), Guimarães (2012), Nascimento (2008), Quintarelli (2005), Gouvêa & Loh (2007), Viera e Garrido (2011), etc

A estilização da atuação do bibliotecário no meio virtual folksonômico por Valentim (2000), Mota e Oliveira (2011), Silva (2004), Passarelli, (2009). A **quinta** seção elenca as categorias de "Dimensões de

Ficção” para literatura de ficção, idealizado por Pejtersen (1978 apud MOREIRA, 2006) para tomar partido das análises de *tags* indexadas/tageadas no overmundo

FOLKSONOMIA NO CONTEXTO DA WEB 2.0

Grande parte das mudanças ocorridas no âmbito da produção de conhecimento é atribuída ao desenvolvimento das ferramentas tecnológica, como os computadores, os quais surgiram inicialmente com o objetivo de armazenar e processar dados. Sobre tal assertiva Johnson (2001, p.110) ressalta

Os computadores começaram como trituradores’ de números, mas passaram a maior parte de sua adolescência sob a tirania de texto - todos aqueles conjuntos de comandos e instruções inescrutáveis resplandecendo em monitores de fósforo verde e inscritos em cartões perfurados. Em seu uso original, de fato [...] input inserido com o acionar de teclas, outp diligentemente transportado para impressora ou monitor. Todas as linguagens importantes que a relação entre computador e usuário era [...]: BASIC, COBOL, Unix, DOS, [...] essas experiências textuais mais velhas nos parecem agora sem vida e opaca [...]

Entretanto os avanços tecnológicos aperfeiçoaram os computadores, além disso, surgiu uma nova geração de ferramentas que transformaram o uso dos computadores. Como afirma Johnson (2001, p. 111) “As novas ferramentas vão ser um meio para ver o todo”. Concernente a essa nova geração de ferramentas, o grande marco na evolução tecnológica, nos últimos 50 anos, se deu a partir do surgimento das tecnologias digital que foram consolidadas com o desenvolvimento da internet.

A internet interliga computadores, formando uma grande rede única de comunicação. Nasceu como uma rede reservada, utilizada apenas por pesquisadores do governo-norte americano, para pesquisas e compartilhamento de recurso de informática, sendo restrita, ao serviço e pesquisa desenvolvida no âmbito militar.

[...] a internet começou a nascer no final da década de 1950 a partir de projetos desenvolvidos por agências do Departamento de Defesa Americano, preocupadas como manter a viabilidade das telecomunicações em caso de guerra nuclear. A ideia central desses projetos consistia em interligar centros militares por meio de computadores, de tal forma que a destruição de um deles não impedisse a sobrevivência dos demais bem como a de um centro remoto [...]. (COSTELLA, 2001, p. 231).

De 1950 a 1970 vários estudos foram realizados sobre as redes de computadores. Entretanto, o estudo mais significativo em termos de impacto em comunicação de computadores foi a *AdvancedResearchPro-*

jectsAgency Network (ARPNET), colocado em serviço em 1971. Em 1985, ARPANET foi renomeada para *Internet*. É interessante destacar a afirmativa feita por Haykin e Moher (2008, p. 21, grifo nosso)

O ponto de virada na evolução da Internet ocorreu em 1990 quando Tim Berners-Lee propôs um software de interface de hipermídia para a Internet, o qual ele chamou de World Wide Web (WWW). Em apenas dois anos, a Web foi da não existência para a popularidade mundial, culminando em sua comercialização em 1994.

Em 1991, Tim Berners-Lee já havia construído todas as ferramentas que precisava: “o navegador, o servidor e as primeiras páginas *web*, que escreviam o próprio projeto” (WORLD... 2012). Segundo o texto WORLD... (2012) um resumo, publicado em 06 de agosto de 1991, marca a estreia da *web* como um serviço publicado na internet. A partir de então, um grande volume de informações é alocado na rede mundial de computadores.

Contudo, as páginas de internet eram textos com hiperligações a que se acrescentou som, imagens e vídeos, permitindo o acesso simultâneo a várias páginas; contudo, nem todo internauta era livre para modificar as páginas que acessava, fosse a termos de conteúdo quanto de linkagem, pois só o programador tinha esta liberdade, não permitindo, assim, a interação do internauta com o conteúdo publicado. (AQUINO, 2007).

Denominada como a “segunda geração de serviços online” a *web 2.0* é caracterizada por Tim O’Reilly (2006 apud AQUINO, 2007, p. 6), o criador do termo,

[...] como não tendo limites rígidos, mas sim um núcleo gravitacional, já que é encarada como uma plataforma, na qual os próprios usuários controlam seus dados, que não mais depende de pacotes fechados de *software*, mas de serviços que rodam no *browser*, online.

Com o surgimento da *web 2.0*, novos tipos de aplicações que suportam publicação de materiais, dão espaço a trocas qualitativas realizadas entre pessoas. Primo (2006) aponta que nesta nova fase da *web*, os ambientes são criados com tecnologias que priorizam novas formas de publicação, compartilhamento e organização de informações.

Sendo assim, sobre a *web 2.0* Johnson (2001, p. 163) filosofa

A tarefa de imaginar a informação não vai mais ser apanágio dos sumo-sacerdotes da programação; qualquer pessoa moderadamente à vontade comum computador será capaz de inventar seus próprios espaços-informação e de partilhá-los com amigos ou colegas. A partir desse sistema mais aberto [...]

Considerando as inúmeras possibilidades de comunicação através da *web 2.0*, evidenciam-se as ferramentas chamadas *software de colaboração*, ou ainda, *software social*, definido por Primo (2003, p.12) como sendo “[...] as tecnologias empregadas para a comunicação entre pessoas e grupos na Internet”, visando à comunicação e organização da informação com suporte à coordenação e cooperação.

Historicamente, a primeira implementação de *software colaborativo* surge na década de 70 com a *Electronic Information Exchange System* (EIES)¹ – Sistema Eletrônico de Intercâmbio de Informação. A primeira menção ao termo “social software” encontrada data de 1990. Depois disso, o termo só foi usado em 1992, por Ted Nelson² e Phil Salin³ (Xanadu e AMIX, respectivamente), usado para expor o que o hipertexto representava (ALLEN, 2008 apud NASCIMENTO 2008).

O *software social* tem desempenhado um papel importante na maneira como as pessoas interagem virtualmente, onde os usuários são os consumidores e os produtores de conteúdo online. O suporte dado à interação faz com que pessoas com interesses semelhantes compartilhem diferentes ideias.

Com vista nas inúmeras possibilidades de comunicação através de *software social*, listamos na tabela 1 algumas das ferramentas e dos recursos provenientes dos conceitos dessa nova proposta para *Web 2.0*.

Tabela 1 - Algumas ferramentas e recursos de *software social*

FERRAMENTA	CARACTERÍSTICAS/FUNÇÃO
Mensagem Síncrona	É considerada 2.0 à medida que permite a presença do usuário no ambiente web, permite a colaboração entre os usuários, empregando-lhe um caráter mais dinâmico. Além disso, os softwares já estão agregando suas funcionalidades a serviços na web, onde permitem a co-navegação, o compartilhamento de arquivos, a captura de telas e o compartilhamento e mineração de dados.
Blogs	No início eram ferramentas utilizadas apenas como diários virtuais, onde as pessoas contavam seus relatos, paixões e atividades rotineiras. Após a disponibilização de softwares próprios para a publicação de blogs, como o Blogger e o Groupsoup, estas ferramentas ganharam uma nova usabilidade, passando de diários pessoais a fontes de informações sobre os mais diversos assuntos. Além disso, os blogs permitem a interação com os leitores, nos espaços disponíveis para comentários.

¹ Sistema de conferência por computador acessível por linha telefônica discada ou redes de computadores desenvolvidas em 1974, por Murray Turoff, com a ajuda de sua esposa Roxanne Hiltz, no então Instituto de Tecnologia de Nova Jersey.

² Ted Nelson fundou o Projeto Xanadu em 1960 com o objetivo de criar uma rede de computadores de interface simples, para o desenvolvimento daquilo que, posteriormente, veio a ser conhecido como hipertexto. O trabalho está documentado no seu livro de 1974 *Computer Lib/Dream Machines* e no livro de 1981 *Literary Machines*

³ Phil Salin fundou a American Information Exchange (AMIX) em 1984, uma rede de compra e venda de informações, bens e serviços. Salin foi pioneiro nos conceitos de compra e venda que são agora consideradas padrão de comércio eletrônico-*e-commerce*

Wikis	São ferramentas que possibilitam que os usuários não só leiam os conteúdos publicados, mas, que também participem da inserção de conteúdos em seu ambiente ou na edição de conteúdos já publicado.
Streaming Media	É uma tecnologia de compreensão de dados que permite ao usuário assistir a um vídeo ou ouvir um áudio sem que haja a necessidade de realizar download do arquivo para o seu computador. Dessa forma, facilita o acesso à informação pelo fato de permitir que o usuário utilize a informação mais rapidamente, visto que a transmissão se dá por meio da própria web.
Redes Sociais	As Redes Sociais contemplam a mais promissora tecnologia web 2.0, pois, elas englobam vários recursos web 2.0 em um único sistema. Elas podem ter como finalidade o simples agrupamento de pessoas por afinidades pessoais ou profissionais, ou, ainda, para a troca de recursos web, como arquivos de dados, fotos, vídeos, áudios ou sites favoritos.
Tagging	Tagging, também conhecida como classificação colaborativa ou, ainda, por Folksonomia, do inglês folksonomy, é o uso de etiquetas ou palavras-chave para a representação de conteúdos na web, sejam imagens, áudios, vídeos, ou páginas web.

Fonte: Adaptado de Rufino (2010)

Tais ferramentas e recursos ganharam grande destaque em meio aos mais variados público, no caso do *tagging*, observa-se que na *Web 2.0* esta forma colaborativa de organizar conteúdos eletronicamente, está ganhando cada vez mais popularidade. É neste contexto que o *tagging*/Folksonomia surge como alternativa de representação da informação no momento em que permite a qualquer usuário da web representar e recuperar informações através de etiquetas criadas livremente e com base nos significados dos dados etiquetados.

Etimologia, nomenclatura e aspectos conceituais da Folksonomia

Em consonância com a literatura fichada, a palavra Folksonomia é uma adaptação da palavra inglesa *folksonomy*, resultante da junção dos termos *folks* – povo/pessoas – e *taxonomy* – taxonomia, estudo e classificação sistemática. Em síntese pode ser traduzida como “classificação feita pelo povo” ou, ainda, “classificação feita por pessoas”. O Arquiteto de Informação Thomas Vander Wal a utilizou em um fórum de discussão no ano de 2004.

Wal (2007 apud NASCIMENTO, 2008, p.31) conta que a ideia não era nova:

Eu sou fã de sistemas de rotulagem e etiquetagem desde o fim dos anos 80 depois de ver um colega de trabalho fazer mágica com o *Lotus Magellan* [...]. Nos anos 90, as pessoas podiam adicionar palavras-chave em documentos e objetos que eles enviavam para as bibliotecas de fóruns da *Compuserve* e o operador do sistema tentava manter as palavras enquanto adicionava termos relevantes de um vocabulário controlado. Entre 1999 e 2000 serviços de etiquetagem, como o Bitzi

vieram para a web e deram a oportunidade dos usuários contribuírem com etiquetas e descrições.

Sobre esta ótica a Folksonomia é apresentada como uma classificação que tem o objetivo de facilitar a recuperação da informação, isto pelo fato de ser realizada pelo próprio usuário, o qual assume papel significativo na web atual. Tal afirmação toma consistência com as palavras de Silva e Blattmann (2007) quando afirmam que:

A finalidade da Folksonomia seria ordenar o caos existente na web. Embora a sua característica de liberdade para classificar aponte para a idéia de uma falta de estrutura organizacional, o resultado para quem pesquisa é uma maior facilidade para encontrar termos que as demais linguagens de indexação não conseguem acompanhar em suas tabelas hierárquicas. (SILVA; BLATTMANN, 2007, p. 207).

O sítio *Del.icio.us* foi o primeiro a utilizar a Folksonomia como recurso para classificação de favoritos, seguido pelo *Flickr*, sítio de hospedagem de imagens, e por outros que se destacam atualmente como sites de vídeos como o *You Tube*, de músicas como é *Last.fm*, de compras como o *Amazon* ou *Submarino*, entre outros; vários segmentos como blogs, sites de relacionamento intranets empresariais como o *Digg* e o *Technorati*, etc. (CALDAS; MOREIRA, 2009; WAL, 2006 apud BROSCH, 2010).



Fonte: Imagem disponível em: <http://www.doisac.com/blog/>

É notório que o termo “Folksonomia” seja utilizado por diversos profissionais das áreas de Ciência da Informação e Biblioteconomia, entretanto é frequente o uso de outras denominações, por não haver um consenso sobre o significado de Folksonomia (CATARINO; BAPTISTA, 2007). Moura (apud CARVALHO; LUCAS; GONÇALVES, 2010, p. 79) fala do uso de termos como “*ethnoclassification*, *collaborative tagging*, social

bookmarking, social tagging, narrow folksonomy, taxonomia popular, classificação social, entre outros”.

Para Wal (2007 apud GUIMARÃES 2012, p. 20), Folksonomia é o “resultado da etiquetagem livre e pessoal de informação e objetos (qualquer coisa com uma URL) para sua recuperação. A etiquetagem é feita num ambiente social (usualmente compartilhado e aberto a outros)”. Não obstante, Catarino e Baptista (2007) entendem o termo da mesma forma: “é o resultado da etiquetagem dos recursos da web num ambiente social, (compartilhado e aberto a outros) pelos próprios usuários da informação visando a sua recuperação”.

Em consonância com os aspectos conceituais dos autores supracitados, a Folksonomia é mostrada como o produto da etiquetagem, mas outros autores entendem a Folksonomia como processo. O quadro 2, elaborada por Catarino e Baptista (2009, p. 50-51, adaptação nossa), mostra diversas definições para Folksonomia, nas duas abordagens.

Quadro 1: Concepções de Folksonomia (produto X processo)

AUTORES	FOLKSONOMIA COMO PRODUTO	AUTORES	FOLKSONOMIA COMO PROCESSO
WAL (2006)	Folksonomia é o resultado da atribuição livre e pessoal de etiquetas (tagging) a informações ou objetos (qualquer coisa com URL), visando à sua recuperação.	RUSSEL (2005)	As Folksonomias têm propiciado a possibilidade de criar desordenadamente, em texto livre, metadados atribuídos pelos usuários para recursos existentes (livros, imagens, URLs, etc).
LUND ET AL. (2005)	Folksonomia se refere a um vocabulário, ou lista de termos, que surge da sobreposição de etiquetas definidas por vários usuários ao marcar as suas hiperligações favoritas, ou seja, seus marcadores para posterior recuperação.	GUY E TONKIN (2006)	Folksonomia é um tipo de sistema de classificação distribuída, e normalmente criada por um grupo de indivíduos, tipicamente os usuários do recurso.
MATHES (2004)	Folksonomia é um conjunto de termos que um grupo de usuários utilizou para etiquetar os conteúdos de recursos digitais da Web.	OHMUKAI, HAMASAKI E TAKEDA (2006).	Trata-se de um sistema que administra etiquetas atribuídas pelos usuários aos recursos por eles indexados, compartilhando-as com outros usuários e também disponibilizando informações de outros recursos disponíveis na Web que foram indexados da mesma forma.

TRANT (2006A, 2006B)	Folksonomia é o resultado de um sistema de classificação socialmente construído, ou, coleção de conceitos expressos num sistema de classificação desenvolvido de forma cooperativa. Folksonomia é um conjunto informal e orgânico de terminologia relacionada.	QUINTARELLI (2005)	Uma nova abordagem emergente para a classificação distribuída de recursos digitais.
STURTZ (2006)	Folksonomia é um conjunto de etiquetas – com uma ou mais palavras-chave – que os usuários de um sistema compartilhado de gestão de conteúdos na Web aplicam a recursos individuais a fim de agrupá-los ou classificá-los para posterior recuperação.	HAMMOND ET AL. (2005)	Uma classificação não estruturada feita pelos próprios usuários dos recursos digitais.
		VALONGUEIRO (2006)	Um novo paradigma de classificação, pois respeita as diferenças culturais e características pessoais de quem utilizou e classificou determinada informação.

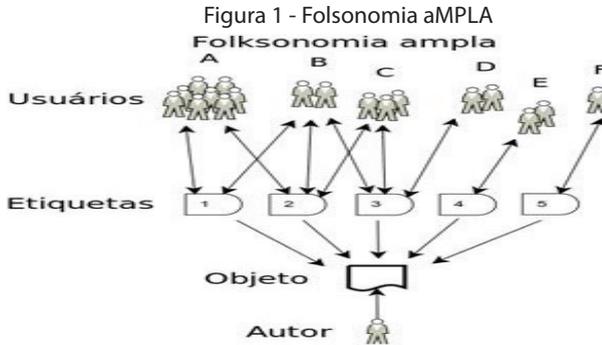
Fonte: (Catarino e Baptista, 2009, p. 50-51)

Diante do exposto no quadro acima, a intenção desta pesquisa é de concordar com as percepções que consideram a Folksonomia como resultado do processo de etiquetagem.

Tipologias e Estrutura de sistemas baseado em Folksonomia

No que diz respeito à tipologia da Folksonomia, Quintarelli (2005 apud AMARAL; AQUINO, 2008) distingue Folksonomias largas de Folksonomias estreitas.

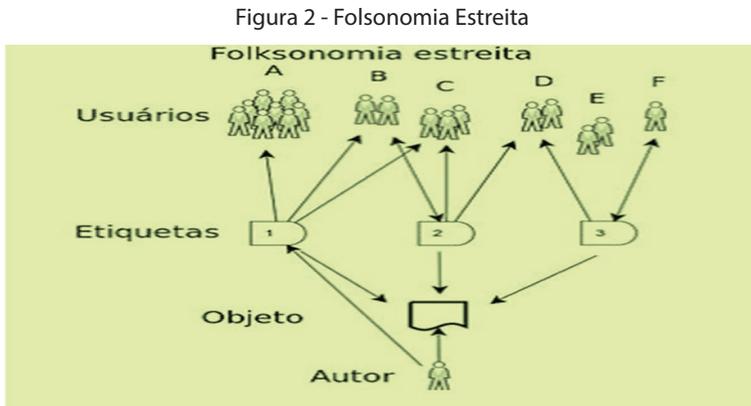
Segundo a autora, nas Folksonomias amplas, muitas pessoas atribuem etiquetas a um mesmo objeto; já nas Folksonomias estreitas, apenas o criador do objeto, ou um grupo restrito de usuários, atribui as etiquetas. As Folksonomias largas são ilustradas na figura 5 por Wal da seguinte maneira:



Fonte: Adaptado de Wal (2005a apud GUIMARÃES, 2012)

Nas Folksonomias amplas, o autor disponibiliza o objeto e/ou informação no ambiente web, e este objeto é etiquetado por grupos de usuários com o mesmo vocabulário, aqui representado pelas letras do alfabeto. Estes grupos conseguem recuperar a informação pelas etiquetas - representadas pelos números - que eles atribuíram a partir de seu vocabulário. É o que acontece nos sistemas de *social bookmarking*, como o *Del.ici.ous*.

Já nas Folksonomias estreitas, um grupo menor de pessoas, ou às vezes somente uma pessoa, coloca etiquetas em um objeto, como no caso do site em pesquisa, o *Overmundo* e o site do *YouTube*. O usuário que postou o conteúdo pode permitir que algumas pessoas também adicionem etiquetas a seus objetos, como no caso do *Flickr*. Dessa forma, os objetos receberão uma quantidade bem menor de etiquetas como ilustrado na figura 2:



Fonte: Adaptado de Wal (2005a apud GUIMARÃES, 2012)

Um sistema que usa a Folksonomia deve ser baseado em *tags*, mas não apenas isso. Se fôssemos determinar regras para identificá-lo, poderíamos considerar que ele deve ter os seguintes componentes, como mostra o quadro 2 abaixo:

Quadro 2 - Componentes de um sistema baseado em Folksonomia

OBJETO	Podemos considerar este objeto como uma URL que recebe as tags (Identificador Universal de Recursos). Ela pode representar um site, uma página específica, um arquivo de texto, uma imagem, um vídeo, uma música ou qualquer outra coisa que se possa acessar via Internet ou que tenha uma representação na rede.
TAGS	São palavras, siglas ou qualquer código pessoal, determinado livremente pelo usuário, de acordo com a sua conveniência. Um objeto pode receber um número ilimitado de tags.
USUÁRIOS	São os agentes do sistema, que atribuem tags aos objetos. A princípio, sua ação é de interesse pessoal, mas desta ação emergem resultados interessantes. Quando vários usuários usam a mesma tag, por exemplo, isso que permite utilizar algoritmos de recomendação, modelo de usuário e filtragem colaborativa em cima delas.

Fonte: adaptado de Lacerda e Valente (2007)

Stakoviak (2011) comunga do mesmo pensamento, quando chama atenção para os três pivôs que compõe processo de categorização de informação, conforme apresentado na Figura 3.

Figura 3 - Os três pivôs da Folksonomia



Fonte: Adaptado de Pereira (2008)

De acordo com a Figura 3, o processo de categorização da informação se dá a partir do **usuário** que realiza a categorização; do **objeto**, que é categorizado; e das **tags**, que categorizam o **objeto**. Sendo assim, o **usuário** atribui um conjunto de **tags** selecionadas livremente conforme sua preferência para descrever, organizar ou classificar o **objeto** em questão.

Vantagens e desvantagens da Folksonomia

Algumas características levantadas na literatura podem ser consideradas vantagens ou desvantagens do uso de sistemas baseados na Folksonomia.

Verificou-se que Catarino e Batista (2007) ao realizarem um estudo sobre o uso da Folksonomia, constataram algumas vantagens e desvantagens. Assim temos o seguinte quadro:

Quadro 3 - Vantagens e desvantagens da Folksonomia

VANTAGENS	DESvantagens
O cunho colaborativo e social	Falta de parâmetro na criação da tags.
A possibilidade de formar, automaticamente, comunidades em torno de assuntos de interesse.	A falta de um controle do vocabulário, que é resultado da característica de liberdade na classificação dos conteúdos.
A liberdade de expressão, que possibilita abarcar todas as formas de ver um mesmo conteúdo, respeitando as diferenças culturais, interpretativas, etc.	Pouca precisão na recuperação da informação, causada pela liberdade de atribuição de etiquetas, pois um mesmo termo pode ter significados diversos para os vários usuários que atribuíram às etiquetas.
O fato de todos os recursos etiquetados estarem disponíveis na Web e, portanto, acessíveis de qualquer computador que esteja ligado à internet.	Imprecisão nos termos utilizados.
A possibilidade de criar uma biblioteca de informação sobre artigos e/ou textos acadêmicos, que também estarão acessíveis em qualquer lugar, não sendo necessário copiar pastas de um computador para outro.	Polissemia e sinonímia.

Fonte: Adaptado de Catarino e Batista (2007)

Há também, Gouvêa & Loh (2007 apud GUIMARÃES, 2012) que corroboram o rol das vantagens do uso da Folksonomia, acrescentando algumas que se diferenciam das autoras supracitadas.

- Pode ser utilizada como filtro de recuperação de documentos e descoberta de conhecimento;
- Possibilita que conteúdos secundários sejam definidos; enquanto taxonomias falam do assunto central ou principal;

- Cria uma estrutura menos rígida e mais flexível, ao contrário das taxonomias, que são limitadas à hierarquia;
- Possuem uma capacidade de se adaptar muito rapidamente às mudanças no vocabulário e às necessidades dos usuários

Pode-se constatar que quanto às suas vantagens e desvantagens a Folksonomia ainda é recente para o usuário que está aprendendo a usufruir de seus benefícios, a mesma *tag* pode levar o usuário a conteúdos que nunca estariam antes associados. Por outro lado, a individualidade de cada usuário influencia a recuperação do conteúdo, pois ao atribuir uma etiqueta a maioria dos usuários está pensando somente em si, ou no grupo com quem se relaciona.

As desvantagens apontadas não desaparecem em curto prazo, pois passa por uma questão de educação do usuário. Entretanto, a Folksonomia está promovendo, embora a passos lentos, a transformação no modo de uso da mesma.

Etiqueta e etiquetagem

Baseada na livre organização, a Folksonomia traz um novo tipo de link, a *tag*, criada pelos próprios usuários que indexam os itens de acordo com seus interesses e necessidades na web, e assim, de forma coletiva representam, organizam e recuperam os dados na rede.

Esse processo de indexação é feito através da etiquetagem (em inglês, *tagging*) em que as próprias pessoas, no caso, os usuários da informação, classificam os objetos ou informações. Golder e Huberman (2005 apud GUIMARÃES, 2012, tradução do autor) explicam que:

Marcar conteúdo com termos descritivos, também chamados palavras-chave ou etiquetas, é uma forma comum de organizar conteúdo para futura navegação, filtragem ou busca. Apesar disso, organizar conteúdo eletrônico dessa forma não é algo novo, uma forma colaborativa desse processo que vem sendo chamada de etiquetagem pelos seus proponentes, está ganhando popularidade na web.

Historicamente, a “[...] etiquetagem de recursos por parte dos usuários é anterior à Folksonomia, e que já no início dos anos 1990 as pessoas podiam adicionar palavras-chave aos documentos enviados ao fórum *Compuserve*” (GUIMARÃES, 2012, p.23). Para o autor, o valor da etiquetagem aparece quando o sistema de etiquetagem traz um objeto comum etiquetado e a identidade de quem o etiquetou. É o que começa a acontecer em 2003, quando o sítio do social *bookmarking Del.ici.ous* é criado.

Nesse sentido, o termo “Etiqueta” (*tag*) refere-se à palavra-chave selecionada pelo usuário para representar o recurso que este está a utilizar. Já o termo “Etiquetagem” (*tagging*) será usado para representar tanto o produto da atividade do usuário quanto à ação de etiquetar recursos propriamente dita.

Segundo N. Junior (2009) em conjunto com as *tags* surgiu à expressão “nuvem de *tags*” que se trata da junção de várias *tags*; essas nuvens são utilizadas como métodos de “taxonomização” no ambiente *web* para hierarquizar estes temas organizando-os em ordem alfabética. Conforme podemos observar na figura 8

Figura 8 - Nuvens de Tags (Tag Cloud)

Ação cultural Arquitetura da informação Arquivamento da web Artigo científico
BD. Sociedade da informação. **Bibliotecário** **Biblioteca** Biblioteca escolar
Bibliotecas Bibliotecas escolares **Bibliotecas** **Públicas** **Biblioteconomia**
Blogs Brasil **Ciência da Informação** cinema **Comunicação científica**
Doenças ocupacionais Estratégias de preservação Fotografia gestão da
informação Gestão do conhecimento Imagem Incentivo à Leitura
Indexação Informação na web Information Literacy Information representation Information retrieval

Fonte: <http://rabci.org/rabci/>⁴.

Consolidando tal perspectiva, Vieira e Garrido (2011) assegura que “O bom funcionamento de um sistema baseado em Folksonomia depende diretamente da indexação promovida por meio de *Clustering*⁵” e mais

O *tag clustering*[...] viabiliza relacionar as *tags* que são geradas automaticamente. Os *clusters* são baseados em URLs comuns, e nos mostram os itens portadores da mesma tag. O modo de navegação é por meio da “nuvem de *tags*” (*tag cloud*), que apresenta um modo de visualização para facilitar a recuperação [...]. A nuvem de *tags* pode indicar a popularidade de uma *tag* ou a frequência da palavra nos documentos da coleção.

Diante disso, pode-se verificar que “A implicação é que, se um grande conjunto de usuários etiquetarem fontes o suficiente, as informações podem ser agregadas para atingir uma maior estabilidade, confiabilidade e consenso” (VIERA; GARRIDO 2011).

⁴ Repositório acadêmico de Biblioteconomia e Ciência da Informação.

⁵ O *Clustering* é uma abordagem da Ciência da Computação para classificação de documentos em uma coleção, baseada nos conteúdos da coleção. *Clustering* explora os relacionamentos dos documentos da coleção, informação semanticamente próxima, pretendendo dividir uma coleção de documentos em grupos (*clusters*).

No caso das Folksonomias, a visualização das etiquetas selecionadas torna-se fundamental para ir ao encontro da função de navegabilidade⁶ (SVENONIUS, 2000; apud VIEIRA; GARRIDO 2011), uma vez que alguns usuários durante a busca por informação, nem sempre sabem exatamente o que buscar ou são incapazes de articular o objeto de sua pesquisa. Ainda assim, de qualquer modo, são capazes de reconhecer imediatamente quando o encontram.

Folksonomia como representação do conhecimento

A despeito de todos os prismas sob o qual vimos a Folksonomia, ainda nos resta averiguá-la como representação do conhecimento. Para tanto, nos subsidiamos em um artigo publicado por Brandt e Medeiros (2010) que tem como objetivo estudar a Folksonomia sob a óptica da representação do conhecimento.

Dentro dessa conjuntura podemos constatar que para que a Folksonomia seja tratada como esquema de representação do conhecimento é imprescindível que façamos uma analogia dos seus aspectos positivos e negativos. E um dos primeiros pontos elencados pelas autoras está no

[...] fato de a Folksonomia ser construída a partir de dados obtidos dos próprios usuários é algo positivo no sentido da garantia do uso, ou seja, o termo usado para representar o documento será o mesmo usado para recuperá-lo posteriormente, por determinado usuário. Neste caso, a garantia de uso do termo na literatura (garantia literária), parece não importar muito, já que a Folksonomia é construída *a posteriori* - não há uma etapa de análise dos documentos do domínio para então se coletar termos e criar posteriormente a base para o esquema de representação, como no caso dos tesouros, por exemplo. (BRANDT; MEDEIROS, 2010, p.118)

Percebe-se que apesar das diferenças nos processos da Folksonomia, ela cumpre os objetivos de um processo de indexação na representação de um documento, sob a visão do usuário, com vista em sua participação. Nos processos de indexação convencional, o usuário não é participante ativo, mas o pressuposto da garantia literária⁷ faz com que o mesmo não seja esquecido. Na Folksonomia, é o usuário que assegura esta garantia, além de ser o responsável pela representação do documento, contribuindo para a facilidade ou não de sua recuperação.

⁶ A navegabilidade é um item importante em um site, pois ela se constitui na organização da informação para que o usuário navegue de forma intuitiva e saiba onde está de onde veio e quais são as suas possibilidades futuras de acesso.

⁷ A garantia literária ocorre quando um vocabulário controlado é capaz de representar conceitos que realmente ocorrem na literatura do assunto.

Levando-se em consideração o requisito de qual unidade deva ser organizada, como acontece nos esquemas de representação tradicionais, as autoras levam em consideração a discussão feita por Broughton *et al.* (2005)⁸ que defendem nas abordagens cognitivas e orientadas a usuários, que a unidade organizável é a estrutura cognitiva, o individual. Portanto, através desse pressuposto a Folksonomia poderia certamente ser considerada uma estrutura de representação do conhecimento, pois será tão mais precisa quanto mais restrito e específico for o domínio representado; as autoras mostram tal raciocínio com a seguinte afirmativa “Representar o mundo cognitivo do usuário é algo que nenhum outro tipo de representação pode fazer tão bem quanto à Folksonomia, ninguém melhor que o próprio usuário para representar seu mundo cognitivo” (BRANDT; MEDEIROS, 2010, p119).

Compreendemos a partir desse pensamento que a grande proficiência do mundo cognitivo do usuário é representada à medida que mais e mais usuários colaborem com suas *tags* em um serviço que vai gerar uma Folksonomia.

Outro aspecto que demonstra a Folksonomia como representação do conhecimento é que ela

[...] tem a vantagem de ser um esquema *botton-up*⁹, em que os termos são gerados a partir do mais específico, não passando pela fase da categorização. Além disso, apesar dos termos serem gerados pelos próprios usuários, estes o faz de uma forma um pouco mais independente que um determinado grupo social escolhido para organizar um domínio (BRANDT; MEDEIROS, 2010, p119).

Entende-se, por conseguinte que a Folksonomia, comparada aos outros esquemas de representação do conhecimento, é gerada de forma inversa: primeiro se classificam os objetos informacionais, e, posteriormente, surge uma *Folksonomia*, representada visualmente pela *tag cloud* (nuvem de *tags*). Coisa que não acontece com os tesouros ou os esquemas de classificação que são usados para a escolha dos termos que irão descrever o conteúdo de um documento que já existe.

A partir dessa análise textual podemos ultimar que a Folksonomia se enquadra nos esquemas de representação, no sentido de ser vista como ferramenta semântica. Todavia devido ao fato dessa modalidade não possuir uma linguagem apropriada no que tange a um vocabulário definido, não pode ser vista e nem ter a finalidade científica ou técnica,

8 BROUGHTON, V.; HANSSON, J.; HJØRLAND, B.; LÓPEZHUERTAS, M.J. *Knowledge Organization*. Report of working group. In: LIS-education in Europe. Working Seminar, at the Royal School of Library and Information Science.

9 *bottonup* (de baixo para cima), onde as regras não são definidas por profissionais e sim por usuários.

porém é um recurso cada vez mais utilizado devido a fácil identificação dos objetos em meio digital (ALBUQUERQUE; MESQUITA; COSTA, 2010).

LITERATURA FICCIONAL E OS ESQUEMAS DE CATEGORIAS DE INDEXAÇÃO PROPOSTAS POR PEJTERSEN (1978 apud MOREIRA, 2006) RELACIONANDO AMBIENTE FISICO E DIGITAL

É pertinente lembrar que esta pesquisa enfoca analisar os critérios de indexação/etiquetagem da seção banco de produtos culturais digitais do site Overmundo, dentre os quais podemos dotar de mais relevância para o nosso trabalho os referidos como literatura, a qual se enquadra na seguinte concepção: “conjunto das obras literárias de um agregado social, [...] ou referidas a determinado assunto como: literatura científica, literatura infantil, literatura de cordel, literatura de ficção, etc.” (MICHAELIS ONLINE). Para efeito de pesquisa, nos detivemos somente a literatura de ficção.

Sabemos que em uma biblioteca, a indexação da literatura ficcional, ou seja, de textos literários, se faz necessária devido à demanda de indagações por parte de seu público alvo que anseia por saber “de que se trata” à obra. Portanto, esta obra deve ser muito bem descrita por um processo de indexação, que é expresso na literatura, segundo Naves (1996, p. 215), como a:

[...] ação de identificar e descrever um documento de acordo com seu assunto [...]. Durante a indexação, os conceitos são extraídos dos documentos através de um processo de análise e, então, traduzidos para os termos de instrumentos de indexação (tais como tesouros, listas de cabeçalhos de assunto, esquemas de classificação, etc.).

A análise de assunto da literatura ficcional é diferente da análise de qualquer outro tipo de literatura, como por exemplo, a literatura científica, devido às características específicas do processo de leitura de um texto literário; onde o indexador encontra um texto carregado de emoção, oposto às áridas linhas que desenhavam a literatura científica ou técnica. Então, Como identificar e descrever sentimentos, conflitos, sensações e sementes da imaginação? Sobre esta questão Lancaster (1993, p.194) expõe:

É possível encontrar alguns procedimentos semelhantes ao analisar um texto literário e um texto técnico, mas a análise de assunto da literatura ficcional [...] demanda um grau de dificuldade maior para o indexador do que outros tipos de publicações.

O texto literário contém e usa mais do que mera informação. O vocabulário é escolhido pela sua sonoridade, leveza e simbolismo. Não existe

qualquer rigor na padronização de vocabulário ou estrutura, é um objeto artístico. Por isso a dificuldade em indexar a literatura ficcional. Mas uma coisa é certa na literatura ficcional, a abordagem temática permite resgatar informações que representem segundo Lancaster (1993):

- o tema ou temas centrais;
- o que a obra pode exemplificar, eventualmente;
- o ambiente em que ela se situa.

A partir desse paradoxo e buscando aprimorar nossa pesquisa, buscamos alguns autores que revelem possibilidades que poderiam ser adaptadas para o tratamento temático do texto literário, o que também incluiria a literatura ficcional, pois a identificação e a seleção de conceitos da mesma requerem o uso de um esquema de categorias mais específicas.

Em nossa revisão de literatura, para Nielsen (1997, p. 176 apud MOREIRA, 2006, p. 42) existem quatro categorias de dados fundamentais relacionadas à literatura ficcional:

1. Personagens (incluindo o narrador);
2. Eventos;
3. Espaços;
4. Tempos.

Entendemos que para Nielsen (1997, p. 176 apud MOREIRA, 2006, p.42), as quatro categorias são relevantes, entretanto, os elementos identificados apenas representam dados objetivos e os aspectos relacionados, tais como: desenvolvimento psicológico do personagem principal; aspectos sociais; aspectos políticos etc., representados por dados objetivos, podem conduzir ao erro. A indexação de textos literários utiliza, geralmente, quatro facetas: Quem (autor e personagens); *Qual* (cenário ação e esquemas); *Onde* (lugares); *Quando* (tempo). A faceta *Como* é relegada ao segundo plano. O autor tem razão ao defender o uso da faceta *Como*. Afinal, identificar na literatura ficcional, personagens, cenário, tempo e espaço é meramente descrição e não análise. Talvez pudéssemos dizer que esta seria a primeira etapa da análise de assunto ficcional. A utilização da faceta *Como* poderia permitir analisar o texto literário ficcional, determinando gênero, estilo, narração e discurso.

Outro esquema de categoria, o qual será de fundamental importância para este trabalho é a categorização denominada *Dimensões da Ficção*, considerado por Lancaster (1993, p. 193) o “método mais aprimorado para a indexação de literatura de ficção”, foi descrito na Dinamarca, pela bibliotecária Annelise Mark Pejtersen. Ela realizou entrevistas de referên-

cia com cerca de trezentos leitores de ficção. Ao analisar as entrevistas, detectou o que ela chamou de “*Dimensões da Ficção*”, caracterizando as demandas dos leitores, enumeradas abaixo (PEJTERSEN, 1978, p. 8 apud MOREIRA, 2006, p. 44):

Quadro 5: Dimensões da ficção idealizado por Perjtersen

Dimensão 1: Assunto (A história é sobre o quê?)	a) Ação e curso dos acontecimentos; b) Descrição e desenvolvimento psicológico; c) Relações sociais.
Dimensão 2: Personagem (a história é sobre quem?)	a) pessoas; b) animais, etc.
Dimensão 3: Cenário (Qual o tempo e espaço escolhido como cenário pelo autor?)	a) Tempo: passado, presente, futuro; b) Local: geográfico, social, profissional.
Dimensão 4: Intenção do autor (Quais as ideias e emoções que o autor quer comunicar com seus leitores?)	a) Experiência emocional; b) Cognição e informação.
Dimensão 5: Acessibilidade (Qual o nível de comunicação?)	a) Capacidade de leitura (“legibilidade”); b) Características físicas e forma literária.

Fonte: adaptado de Moreira (2006)

Observa-se que cada dimensão foi organizada de acordo com um conjunto próprio de critérios, sendo subdivididas em quatro categorias independentes.

Segundo Pejtersen (1978, apud Moreira 2006, p. 39), os estágios do processo de indexação, na literatura ficcional, podem ser descritos da seguinte forma:

1. identificação dos aspectos pertinentes para cada ponto de vista;
2. julgamento da predominância dos aspectos identificados;
3. seleção dos aspectos que atendam às necessidades do usuário, ao desejar um documento com um determinado elemento.

É importante salientar que não existe uma ordem lógica para o estabelecimento dos pontos de vista. Cada dimensão não depende uma da outra, assim como suas subdivisões, de forma que, qualquer assunto pode ser expresso. Elas representam uma combinação de algumas funções e aspectos da ficção. (PEJTERSEN, 1979, p. 255 apud MOREIRA, 2006, p.45).

No que diz respeito ao ambiente virtual essa problemática não difere muito, pelo contrário se acentua ainda mais, por meio da mudança

do paradigma do texto linear para o texto em rede, o qual irrompe não apenas um novo tipo de leitor – o navegador, mas também se daria o desaparecimento do autor e de sua autoridade ou, pelo menos, uma reconfiguração de seu papel e *status*. Mais ainda, em contrapartida ao leitor e autor, surgiria uma espécie de “lautor” (*wreader*), que reuniria em si tanto o consumidor quanto o produtor de textos (BELLEI, 2002, p. 70-1).

Adendo à afirmação anterior Brandt (2009, p. 60) sugere que

[...] para recursos informacionais em que prevalece o caráter subjetivo, a proposta de indexação colaborativa ou democrática surge como alternativa para ampliação e melhora do acesso a esses conteúdos.

A Folksonomia por ser um processo feito através da etiquetagem, ou seja, atribuição de etiquetas, num processo livre e espontâneo no qual o usuário utiliza os termos que melhor preferir para uma posterior recuperação do item digital, no caso do texto literário, cairia bem a calhar como um método alternativo para a indexação da literatura ficcional no ambiente virtual.

No pensamento de unir útil ao agradável, nesta pesquisa associaremos o esquema de categorias idealizado, por Pejtersen (*Dimensões da Ficção*) aos sistemas baseados em etiquetagem usados atualmente. O quadro 6 reproduz a grade proposta por Pejtersen (1978 apud MOREIRA, 2006), que será utilizada no estudo de caso deste trabalho

Quadro 6: grade de classificação proposta por Pejtersen (1978 apud MOREIRA, 2006)

USUÁRIO:				
TÍTULO:				
CATEGORIAS DE PEJTERSEN				
Assunto O QUE?	Personagem QUEM?	Estrutura/ Cenário ONDE?	Intenção do Autor COMO?	Acessibilidade QUAL?
TAGS:				

Fonte: Moreira (2006) com adaptação da autora

NAVEGANDO ENTRE OS CAMINHOS METODÓLOGICOS

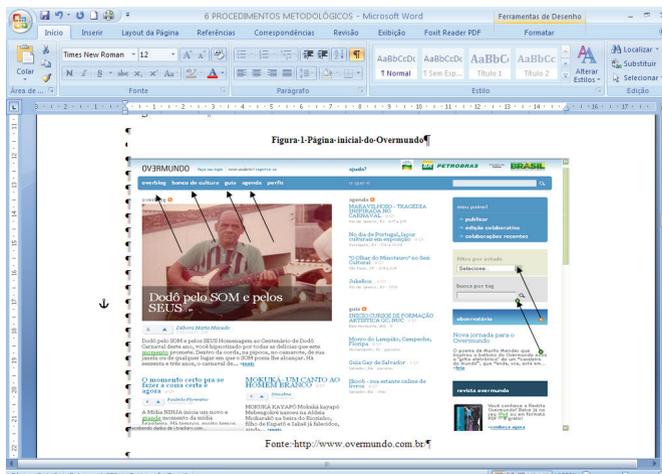
Um dos elementos fundamentais para condução de qualquer trabalho científico é a escolha metodológica, que se constitui na seleção de procedimentos sistemáticos e/ou estratégias de pesquisa para descrição e explicação de uma determinada situação de estudo (MUELLER, 2007).

OVERMUNDO: fazendo um tour virtual no ambiente da pesquisa

É um *website* colaborativo voltado para a cultura brasileira, lançado em março de 2006 pelo Movimento Núcleo de Ideias, formado por Hermano Viana, José Marcelo Zacchi, Ronaldo Lemose Alexandre Youssef, por meio de patrocínio da Petrobras. Hoje, o site é administrado pelo Instituto Overmundo com o objetivo de dar visibilidade na internet à produção cultural brasileira que não é vista na grande mídia. Ele conta com artigos, um guia cultural das cidades brasileiras, uma agenda cultural e um banco de produtos culturais. Qualquer visitante pode criar uma conta e publicar, votar ou sugerir edições do site. O grande desafio proposto pelo Overmundo é a difusão da produção cultural brasileira em toda a sua diversidade.

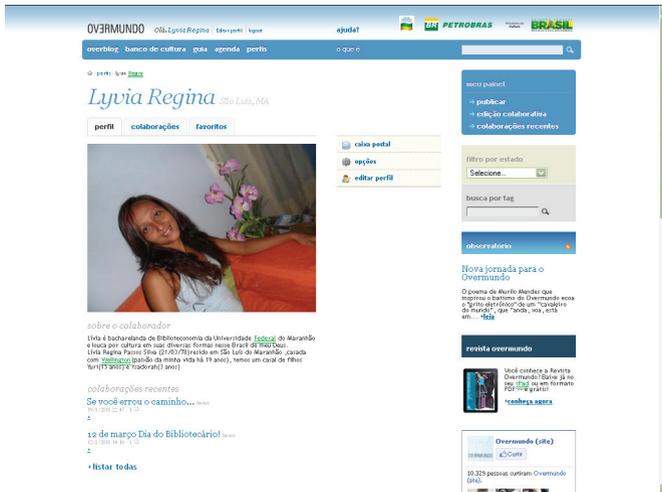
Neste *website* encontramos reportagens, entrevistas, críticas sobre produção cultural brasileira; textos de ficção e não-ficção, poemas, teses, dissertações, monografias, livros, fotos, obras de arte, músicas, discos, filmes, dicas culturais e programação cultural das cidades brasileiras. Tudo organizado em quatro seções principais: *Overblog*, Banco de Cultura, Guia Cultural e Agenda Cultural. Como mostra a figura abaixo:

Figura 4 - Página inicial do Overmundo



A busca por publicações no Overmundo pode ser realizada clicando no campo de filtro por localidade e depois inserindo tags ou nomes de usuários. Ao inserir um termo para busca, a relação sugerida pelo sistema é de tags relacionadas com a utilizada na busca. A relação entre as tags se dá a partir das atribuições feitas pelos usuários no momento da publicação de conteúdos. Para publicar no Overmundo basta registrar-se, fazer o seu login e começar a interagir com os conteúdos e a comunidade do site.

Figura 5 - Página de um usuário cadastrado no Overmundo



Fonte: <http://www.overmundo.com.br/>

A figura 6 mostra a disposição das informações da página de um usuário cadastrado no Overmundo. Após a criação da conta de usuário, é possível postar sua colaboração em uma das seções do site; atribuindo *tags* representativas aos conteúdos. Além disso, é possível ter acesso, ainda, a todas as *tags* utilizadas pelo usuário em suas colaborações. Através das *tags* qualquer colaborador pode acessar seu perfil clicando na tag para o seu nome em uma das colaborações que submeteu ao site. Você pode também navegar pelos perfis de todos os integrantes da comunidade do Overmundo ou procurar por um deles em particular por meio da *tags* perfil.

A página do perfil traz um espaço para fotos do usuário e informações básicas, como nome, cidade e estado, além de indicar desde quando aquele usuário é membro da comunidade do Overmundo. Além disso, a página traz uma lista das colaborações postadas por este usuário, bem como seus comentários e colaboradores favoritos.

No Overmundo existe um espaço onde é possível trocar mensagens com outros usuários dentro do próprio sistema, ou seja, há uma interação dialógica entre os usuários. Para que uma pessoa seja contato de outra dentro do Overmundo é preciso aceitação, dependendo dele à aceitação ou não de mensagens de outras pessoas cadastradas no site.

Figura 6 - Página opção de seção para publicação



Fonte: <http://www.overmundo.com.br/>

No caso de publicação, o colaborador deve clicar em Publicar colaboração, fazer a opção pela seção que mais se adéqua para sua colaboração. Por exemplo, um usuário colaborador deseja fazer uma publicação de uma poesia. Ele deverá clicar em publicar colaboração, fazer a opção pela seção Banco de cultura e em seguida escolher a categoria do produto a ser publicado, que no caso, será a categoria poesia.

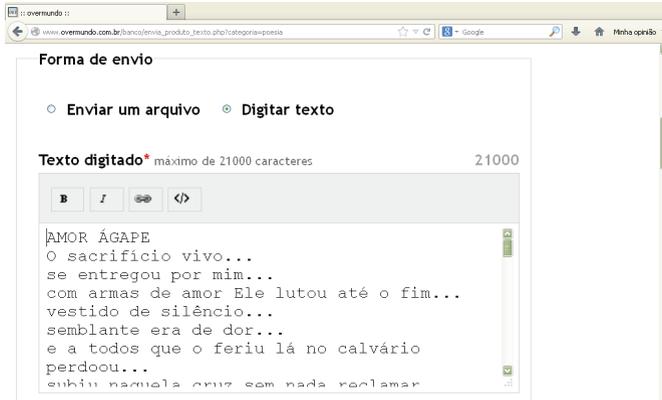
Figura 7 - Página opção de categorias de publicação



Fonte: <http://www.overmundo.com.br/>

Na página de publicação, escolhe uma das opções: enviar um arquivo de texto (para arquivos grandes para download) ou digitar um texto (para leitura direta no post, com o limite de 21 mil caracteres).

Figura 8 - Página de forma de envio da publicação



Fonte: <http://www.overmundo.com.br>

O sistema de indexação/tagagem do conteúdo do Overmundo também é colaborativo. No momento da postagem, você deve escolher as palavras-chave (ou *tags*, como ficaram populares recentemente na internet) que melhor resumem os assuntos nela abordados. No caso da poesia publicada, as *tags* usadas foram Jesus e sacrifício-vivo. E estas *tags* vai servir para *linkar* seu texto com todo o resto do conteúdo do Overmundo. Como podemos observar na figura abaixo:

Figura 9 - Página de criação de tags



Fonte: <http://www.overmundo.com.br/>

O usuário colaborador pode associar sua colaboração às *tags* que quiser, mas, para facilitar a busca, o site aconselha o uso preferencialmente no singular (“filme” e “blog”, em lugar de “filmes” e “blogs”), sem acentos ou cedilha. As *tags* devem conter no mínimo dois caracteres alfanuméricos e virem separadas umas das outras apenas por espaços. Além disso, no caso de expressões ou palavras compostas, use hífen para indicar que se trata de uma única *tag*. Outra opção que o site dá para escolha das *tags* é Nuvens de *Tags* (*Tag Cloud*) expressando os assuntos mais populares do Overmundo.

Figura 10 - Tags da colaboração publicada



Fonte: <http://www.overmundo.com.br/>

Além das *tags* que o usuário colaborador cria, as suas colaborações no Overmundo são também associadas à categoria (Música, Cinema/ Víde, Literatura, Artes Cênicas etc.), ao estado e à cidade escolhidas por você no momento da publicação. Essas são *tags*-padrão que facilitam e ajudam também a consolidar a classificação colaborativa dos conteúdos no site.

Figura 11 - Página de colaboração publicada com mecanismo de votação colaborativa

Fonte: <http://www.overmundo.com.br/>

Depois de publicada, a colaboração do usuário já pode ser votada pela comunidade, que seleciona os destaques que ocupam a primeira página do site, por meio de um mecanismo de votação colaborativo. Por isso, cada colaboração tem um botão de voto disponível à esquerda de cada publicação. Conforme ela recebe mais votos, tem mais destaque; mas perde destaque há quanto mais tempo estiver publicada.

Ainda que outras ações possam ser realizadas no Overmundo, essas descritas acima são as que se referem ao uso da Folksonomia dentro do sistema.

Visto os detalhes do site, o funcionamento do sistema de etiquetagem e suas peculiaridades irá facilitar o entendimento da pesquisa.

Abordagem e tipo de pesquisa

A abordagem aplicada nesta pesquisa foi mista, tanto qualitativa como o estudo qualitativo costuma ser direcionado ao longo do seu desenvolvimento e não busca enumerar ou medir eventos, além de não utilizar ferramentas estatísticas para análise de dados. Conforme Neves (1996) faz parte da pesquisa qualitativa à obtenção de dados descritivos mediante contato direto e interativo do pesquisador com a situação do objeto de estudo.

Já o estudo quantitativo busca a explicação para as causas da pesquisa. A esse respeito, afirma Figueiredo (2004, p. 107-108), que o estudo quantitativo “[...] associa análise estatística à investigação dos significados das relações humanas, privilegiando a melhor compreensão do tema a ser estudado e facilitando desta forma a interpretação dos dados obtidos”.

Enquanto o método quantitativo confirma ou generaliza os resultados e conclusões, o método qualitativo tende a interpretar a apresentação de resultados. Sobre essa questão Minayo (1994, p. 22), explica que “o conjunto de dados quantitativo e qualitativo, [...], não se opõem. Ao contrário, se complementam, pois a realidade abrangida por eles interage dinamicamente, excluindo qualquer dicotomia”.

Nesta pesquisa o estudo quantitativo foi apresentado a partir de três gráficos elaborado pela autora: dois para quantificar as *tags* postadas na primeira e na última publicação e outro para classificar as *tags* postadas pelos usuários por meio da metodologia de análise de assunto para a indexação dos textos literários criada por Pejtersen (1978 apud MOREIRA, 2006).

O estudo qualitativo foi embasado na metodologia de análise de assunto para a indexação dos textos literários criada por Pejtersen (1978 apud MOREIRA, 2006), onde analisamos e interpretamos cada *tag* utilizada pelos usuários estudados no presente trabalho.

Instrumentos da coleta de dados

Os instrumentos que auxiliam o pesquisador em qualquer pesquisa são extensões do próprio pesquisador e têm a função de ajudá-lo a captar situações importantes para depois revê-las e analisá-las (NASCIMENTO, 2008, p. 56). Portanto, o presente trabalho foi desenvolvido a partir de dois elementos importantes para netnografia, quanto a coleta de dados:

- Observação participante¹⁰;
- Entrevista on-line¹¹.

A técnica da observação participante, segundo Vidich e Lyman (2006) indica que a observação pode ser estruturada ou livre. Entretanto nesta pesquisa adotamos a observação livre por esse tipo satisfazer as principais necessidades da pesquisa qualitativa e por não se necessitar de estabelecimento prévio de categorias que somente podem se constituir na análise.

Embora não existam regras fixas do que observar Kozinets (2002 apud NASCIMENTO 2008) a itens que em virtude de serem significativos, foram considerados pelo pesquisador:

- o cenário (Overmundo);
- as pessoas (usuários colaboradores);
- ato (técnicas de tagueamento);
- evento (conjunto das atividades relacionadas).

Outra técnica utilizada pela netnografia para coletar dados é a entrevista on-line. Nesta, opera-se na premissa do processo dialógico entre entrevistador e entrevistado. Por isso a presença constante do pesquisador no campo é necessária para que este conheça profundamente os outros membros e suas formas de comunicar-se com o grupo WARD (1999 apud NASCIMENTO 2008).

Para Christians(2006) a ideia não é fazer um conjunto de perguntas padronizadas ou esperar que o entrevistado traduza os pensamentos do

¹⁰ Dados que o pesquisador obtém ao observar a comunidade, seus membros, interações e significados.

¹¹ Dados que o pesquisador copia diretamente das comunicações por computador efetuadas pelos membros das comunidades on-line.

pesquisador em categorias de respostas, mas que as perguntas sejam um convite para o entrevistado falar longamente (ou não) com suas próprias palavras.

Por isso, os dados coletados foram descritos e classificados de forma que sua análise foi baseada na teoria que compõem o referencial teórico, principalmente no que se refere à Pejtersen (1978 apud MOREIRA, 2006) e suas interpretações ficou a cargo de gráficos estatísticos.

Configuração da pesquisa

Quanto aos objetivos propostos, este estudo de caso seguiu o modelo de aplicação da netnografia a qual, segundo Kozinets (2002 apud NASCIMENTO, 2008 p. 49) trata-se de:

[...] uma descrição escrita resultante do trabalho de campo que estuda culturas e comunidades on-line emergentes, mediadas por computador, ou comunicações baseadas na internet onde trabalho de campo como a descrição textual são metodologicamente conduzidas pelas tradições e técnicas da antropologia cultural.

Dessa forma, o presente estudo foi desenvolvido em três fases:

Quadro 7 - Modelo de aplicação da netnografia

1ª FASE	ENTRADA NO CAMPO	Revisão bibliográfica Identificar os usuários colaboradores que farão parte da pesquisa
2ª FASE	COLETA DE DADOS - NETNOGRAFIA	<u>Observação Participante:</u> Encontrar 2 produtos que tenham sido tageado por 3 usuários colaboradores do Overmundo; Examinar as tags dos produtos encontrados; Classificar as tags examinadas por meio da metodologia de análise de assunto para a indexação dos textos literários criada por Pejtersen (1978 apud MOREIRA, 2006). <u>Entrevista Online:</u> Contato virtual com os usuários colaboradores do Overmundo, a fim de convidá-los a participar da pesquisa; Saber dos usuários colaboradores a motivação para para tageamento de seus produtos.
3ª FASE	GARANTIA DA CONFIABILIDADE DA PESQUISA	

Fonte: Adaptado de nascimento, 2008

Esta pesquisa foi realizada com os usuários que fizeram uso do Overmundo no ano de 2013 (monografia). A partir do cadastro no Overmundo, buscamos uma amostra de 3 usuários com perfis atualizados no período de março a agosto de 2013. Justificou-se esta porção da população da amostra pelo fato de realizarmos a pesquisa em um ambiente dinâmico, por isso julgamos estabelecer um curto espaço de tempo, e trabalhar com usuários que utilizaram o site e atualizaram seus perfis no período estabelecido, pois as constantes mudanças ocorridas no contexto da web 2.0 poderiam dificultar a coleta de dados.

Outra escolha foi a opção por apenas às categorias de literatura de ficção e poesia da seção banco de cultura, já que a seção dispõem de artes visuais, vídeos e músicas.

Após definida a categoria, realizamos a análise netnográfica que foi baseada em dois pontos. Primeiramente, selecionamos 2 (duas) publicação de cada um dos 3 usuários da amostra, usando como critério a primeira e a última publicação. E, em um segundo momento analisamos as *tags* postadas nesses produtos por meio da metodologia de análise de assunto para a indexação dos textos literários criada por Pejtersen (1978 apud MOREIRA, 2006).

Em outro momento da pesquisa netnográfica aplicamos o instrumento de coleta de dados (entrevista online) que foi organizado de tal forma que seu conteúdo foi estruturado baseando-se nas questões da pesquisa.

A entrevista online coletou dados a partir de diálogos pela caixa postal de cada colaborador cadastrado no *Website* Overmundo. Nossa coleta compreendeu o período de 12/03/2013 a 15/03/2013. Para preservar o anonimato dos entrevistados estes foram identificados apenas pela consoante "C", indicando a palavra colaborador e acrescida de números. Este cuidado ajuda o processo de interpretação e análise dos dados. Todos 3 (três) colaboradores escolhidos para participar de nossa pesquisa responderam ao nosso e-mail.

Coletados os dados da entrevista netnográfica, parte-se para a fase de análise dos dados.

ANÁLISE E INTERPRETAÇÃO DE DADOS

As 6 publicações analisadas somaram um total de 47 *tags*, dando um margem de 7,83 *tags* por publicação. A publicação com mais *tags* possuía 10 e a com menor número apenas 6 *tags*.

A seguir são apresentadas as publicações analisadas seguidas das categorias da grade proposta por Pejtersen (1978 apud MOREIRA, 2006) preenchidas pelas *tags* postadas.

Análise de publicação 1

“MISSÃO”

Amar Doar-se

Sem pensar em si Mas sim no próximo.

Respeitar e agradecer A Natureza

Por tudo

Em qualquer tempo. Esse é o meu destino.

Espero conseguir Sabedoria Para essa missão...

Fonte: <http://www.overmundo.com.br/>

Nesta publicação o C1 postou as seguintes *tags*: “amor”, “vida”, “carinho”, “respeito”, “agradecimento”, “poesia”.

As categorias propostas por Pejtersen (1978 apud MOREIRA, 2006) em sua já mencionada grade de metodologia de análise de assunto para a indexação dos textos literários foram preenchidas da seguinte forma:

Quadro 8 - Distribuição das *tags* da publicação 1 classificadas na grade de classificação proposta por Pejtersen (1978 apud MOREIRA, 2006)

USUÁRIO: C1				
TÍTULO: Missão				
CATEGORIAS DE PEJTERSEN				
Assunto O QUE ?	Personagens QUEM?	Estrutura/ Cenário ONDE?	Intenção do Autor COMO?	Acessibilidade QUAL?
respeito agradecimento			amor vida carinho	poesia
TAGS: amor, vida, carinho, respeito, agradecimento, poesia				

Fonte: Moreira (2006) com adaptação da autora

Análise de publicação 2

MEU BATMÓVEL AMARELO

Comprei do Batman, ele anda apertado depois que último filme não rendeu o esperado e ainda por cima, a mulher gato o está processando por tê-la chamado de machorra. Ela embora tenha pensado que suas desconfiças em torno da dupla Batman e Robim tinha algo de estranho. Mas como não é burra, resolveu ficar quieta e sacar uma boa grana do morcego. Lá nos States, isso é uma séria violação do tratado dos direitos de homossexuais, lésbicas, bissexuais e afins. Pinteí a tralha de amarelo, porque não gosto de preto e tenho saudades da Brasília amarela dos Mamonas. De lambuja ele me deu a mulher gato como brinde. Mas, não gosto de gatos, principalmente com gatas com preferências sexuais contrárias aos meus interesses. Então, eu prefiro a Christiane, uma indiazinha que apareceu lá em casa e não foi mais embora, esta que aparece nua deitada na areia do deserto do Saara. Nem estava tão quente, mas ela prefere andar nua sempre que possível. Não apareço na foto porque também estava nu e além da questão do pudor, alguém tinha que fazer a foto que vou anexar no processo contra o morcego trapalhão. Vendi a mulher gato pra Globo. Ela deve estrear em breve no Zorra total. Quanto ao Batmóvel, ferveu no deserto. Isso já me deixou puto porque fui esnobar nas dunas em Santa Catarina e ele pipocou. Fui gozado por um velhote que subiu a toda com um bug a base de fusca. Como se não bastasse, o rádio não funciona e o GPS é só cenográfico. Não serve pra nada. No cinema ele sobe até em parede, mas aqui pipoca em qualquer rampa. Bem! Você deve estar pensando de onde tirei essa história maluca. Eu também. Pra começar eu nunca vi um filme do Batman. Nem fazia parte das minhas preferências quando garoto. Sempre preferi histórias mais adultas. Também nunca estive no deserto do Saara. Só sei que acordei com a Christiane pelada deitada em cima de mim, todo suado e com a calçinha dela enrolada no pescoço. Ok, foi só um sonho. Mas não foi um sonho como qualquer outro. Então levantei e fui pro computador. Tinha que escrever esta história pelo inusitado do sonho nada parecido com os sonhos habituais. Nestes a gente acorda e pronto! Foi só um sonho. Neste aconteceu algo fora do comum. Ou de onde você acha que tirei a foto?

Fonte: <http://www.overmundo.com.br/>

Nesta publicação o C1 postou as seguintes *tags*: "bilheteria", "mulher-gato", "saara", "batmovel", "indiazinha", "pelada", "calcinha", "pescoco", "ficcao".

As categorias propostas por Pejtersen (1978 apud MOREIRA, 2006) em sua já mencionada grade de metodologia de análise de assunto para a indexação dos textos literários foram preenchidas da seguinte forma:

Quadro 9 - Distribuição das tags da publicação 2 classificadas na grade de classificação proposta por Pejtersen (1978 apud MOREIRA, 2006)

USUÁRIO:C1				
TÍTULO: Meu batmóvel amarelo				
CATEGORIAS DE PEJTERSEN				
Assunto O QUE ?	Personagens QUEM?	Estrutura/ Cenário ONDE?	Intenção do Autor COMO?	Acessibilidade QUAL?
batmóvel	mulher-gato batmovel indiazinha	bilheteria saara	pelada calcinha pescoco	ficção
TAGS: bilheteria, mulher-gato, saara, batmovel, indiazinha, pelada, calcinha, pescoco, ficção				

Fonte: Moreira (2006) com adaptação da autora

Análise de publicação 3

ALZIRA E O LOBISOMEM

A cidade estava em pavorosa, era só a lua cheia cair numa noite de seixa-feira que a fera atacava. As cabeças de gado nas fazendas eram devoradas ou simplesmente desapareciam. Poças de sangue se encontravam pelos caminhos. Os rumores cresciam avassaladoramente, haviam relatos impressionantes de que pessoas tinham sido atacadas e escapado milagrosamente da criatura. Ninguém sabia ao certo o nome das supostas vítimas, mas do agressor todos tinham certeza: era um lobisomem.

Os ataques foram se sucedendo de forma misteriosa. Suspeitava-se de todos que poderiam carregar consigo a sina do lobisomem: o sétimo filho de uma linhagem de sete irmãos, os forasteiros e os ciganos, gente muito peluda, de olhar aguçado ou com estranhas cicatrizes dais quais não sabia explicar direito a origem.

O delegado, inteiramente cético em matéria de lobisomem e outras assombrações, tinha suspeitas mais realistas: poderiam ser ladrões de gado, brincadeiras da juventude do lugar ou então, gente que se aproveitava da onda de superstição que tomou conta do lugar para dá fim aos desafetos.

Entre crença e descrença, os boatos se alastravam. Muitos já duvidavam de fato da existência do amaldiçoado, até a fatídica noite em que Alzira foi atacada. Alzira, moça estudada e de muito recato, não era chegada a crendices e outras coisas que ela mesma denominava bobagens dessa gente. Tinha atitude e postura e, sua palavra inspirava credibilidade. Entretanto, o maior de seus atributos era mesmo a sua infundável beleza e formosura. Dona de um corpo escultural, pernas

torneadas, seios fartos e rígidos, de rosto simétrico, ornamentado por um fabuloso par de olhos castanhos e arredondados, e longos cachos negros que realçavam seus ombros alvejadados.

A bela Alzira atraía para si todos os olhares, aticava paixão e desejo em todos os homens da cidade, os solteiros e os impedidos. As mulheres também, umas com respeitável admiração, outras com pura inveja, e outras por razões que não sabiam explicar. Com tantos olhares sobre si, Alzira não poderia passar despercebida ao excomungado, que em sua forma de gente passou a segui-la por todos os lugares. Encantado com a formosura da moça, o maldito teve fome de sua succulenta carne.

A malfada sexta-feira de lua cheia chegou, Alzira voltava sozinha da Igreja, posto que declinou a gentileza de um admirador que se dispôs a acompanhá-la até sua casa, trajava um vestido vermelho de laço, que deixava a amostra um belo par de panturrilhas. Não sei se foi a vermelhidão do vestido ou se foi a pele branca a sobressair por baixo da bainha, ou se foi simplesmente Alzira, por ser Alzira que atraiu o monstro. Primeiro se ouviu de longe um uivo aterrorizante, o sangue de Alzira gelou e ela se pôs aos berros a correr pelos caminhos, atravessou uma vereda perto a um enfarpado onde os arames lhe rasgaram e arrancaram o vestido, deixando-a inteiramente despida.

O lobisomem em seu rastro farejou o vestido e o cheiro de sangue, que vazou de seus muitos arranhões, e foi certo ao encontro de Alzira, que se escondia atrás de um cajueiro. As narinas do monstro exalavam e seus olhos famintos fitaram a presa. Alzira acuada, em choro compulsivo, se prostou deitada ao chão e esperou o bote da criatura.

O lobisomem se aproximou e lambeu Alzira nas partes. A bela gemeu sem saber se de pavor ou de prazer, e olhou para o desalmado, dominada por uma ternura estranha que resultou em um beijo.

Não é que o bicho aos poucos foi tomando forma humana e em ato repentino possuiu carnalmente a moça que de fato era donzela, deixando no local mais uma poça de sangue. Atraídos pelos primeiros gritos de Alzira, os bravos do lugar vieram em seu socorro empunhando tochas, rifles, pistolas e facões. A encontraram desfalecida, nua e ensangüentada, com visíveis sinais da agressão sofrida. Alguns ainda viram a vulto da criatura sumindo por entre a folhagem. Quando lhe perguntaram quem teria feito aquilo, Alzira não pensou duas vezes: "Foi o lobisomem!"

Depois do ocorrido, as moças do lugar temiam a desejada sorte de cruzar com o lobisomem...

Fonte: <http://www.overmundo.com.br/>

Nesta publicação o C2 postou as seguintes tags: "lenda", "alzira", "lobisomen", "fantasia", "humor", "sensualidade", "seducao", "terror", "ficcao".

As categorias propostas por Pejtersen (1978 apud MOREIRA, 2006) em sua já mencionada grade de metodologia de análise de assunto para a indexação dos textos literários foram preenchidas da seguinte forma:

Quadro 10 - Distribuição das tags da publicação 3 classificadas na grade de classificação proposta por Pejtersen (1978 apud MOREIRA, 2006)

USUÁRIO: C2				
TÍTULO: Alzira e o lobisomem				
CATEGORIAS DE PEJTERSEN				
Assunto O QUE ?	Personagens QUEM?	Estrutura/ Cenário ONDE?	Intenção do Autor COMO?	Acessibilidade QUAL?
Lenda	alzira lobisomen		fantasia humor sensualidade seducao	ficção terror
TAGS: lenda, alzira, lobisomen, fantasia, humor, sensualidade, seducao, terror, ficcao				

Fonte: Moreira (2006) com adaptação da autora

Análise de publicação 4

PARAÍSO

A vida lhe tem sido ingrata, ou senão, tem escarnecido do pobre coitado. A sucessão de fatos adversos, o tem condenado a uma espécie de morte moral e social, lenta e fragmentada em outras mortes.

O casamento despencou na rotina e no excesso de falta financeira. Um ódio velado, nascido do fato de que um dia um foi apaixonado pelo outro, foi o fungo que durante os longos vinte anos se nutriu da incompatibilidade do casal, corroendo cada momento da vida em comum, vomitando sobre cada fagulha do amor que um dia os uniu. O matrimônio enfadonho foi sua primeira morte.

A segunda morte foi o divórcio, que ninguém se engane, ainda é pior que o casamento infeliz. A falsa sensação de liberdade se dissolve quando se percebe que o vínculo do matrimônio de fato é eterno, permanece o mesmo desrespeito, as mesmas aporrinhações, os mesmos problemas, a mesma falta de compreensão, de diálogo e companheirismo, só que tudo a distancia. O fato é que a distancia não é só do ex-cônjuge, é de tudo, da vida construída, dos filhos, de si mesmo.

A terceira morte foi o alcoolismo, que já existia em pequenas doses, e agravou em litros, depois do divórcio. Vexames, vômitos, vadiagem, viadagem, vagabundas, verrugas, vícios, verborrêia, vacilos, vazio, veneno, vesgo, vencido, foram alguns dos muitos "V" que se associaram

a palavra Vodga, fazendo de sua vida sem sentido uma verdadeira droga.

O fumo, foi a sua quarta morte. Não tão lenta como muitos imaginam, nem tão pouco menos dolorosa. Os pulmões, lesos e esfolados pelas dezenas de substâncias tóxicas que neles se acumulavam, abriam-se em chagas pútridas. O fôlego se perdia e os degraus que surgiam em seu caminho se multiplicavam, um simples passeio se convertia em penosa jornada.

A quinta morte foi o desemprego que se deu em virtude da terceira e da quarta morte. A perda do ofício resulta no final em perda total da dignidade, do respeito e da condição humana. Traduz-se em um olhar perdido nos dias de fome. No âmago da desesperança mora a escassez do mínimo necessário para sobrevivência.

A sexta morte foi ter sido surpreendido, em meio a quinta morte, pela notícia da doença da filha. Sofreu angustiado a impotência perante o inevitável. A sétima morte foi a morte iminente da filha.

A oitava morte foram todos os outros vícios que se acumularam. A dependência incontrolável a incontáveis substâncias nocivas e proibidas por lei. O demônio entrava por suas veias, controlava sua mente, devorava seus neurônios sobreviventes e consumia seu corpo. Os anos encurtaram e a velhice precoce foi sua nona morte.

Naquela noite, Chegou chapado. Era impossível dizer se só cheirou, se foi só pico, ou se apenas, como um lorde acendeu um cachimbo de latão, com pedras brancas, e fumou ao lado de uma família de sem-tetos embaixo de algum viaduto.

O barracão, improvisado com placas de zinco e madeira, sobras roubadas de outdoors, ainda guardavam restos dos anúncios que antes emolduravam (anúncios que vendiam carros e apartamentos caros, e a juventude estampada das vitrines), era quente, e no verão, depois de ter sido submetido a um sol escaldante, um mornaço insuportável tomava conta de todo o ambiente durante a noite. Acostumado ao calor já quase o não sentia. Ligou o rádio e foi saudado pela voz de Nelson Gonçalves. Inspirado abriu a garrafa de vodga (que nunca lhe faltou, mesmo depois de ter perdido o emprego), acendeu um cigarro e sentou-se na beira do colchão, que também era uma cama improvisada no chão batido, e fechou os olhos.

Pouco depois para sua surpresa ouvia a voz da filha lhe chamando lá fora. Era uma noite bem mais quente que todas as outras, ainda assim sentiu uma brisa suave entrar quando abriu a porta.

Ela estava linda, bem vestida e cheia de vida, nem parecia que tinha sofrido tanto naquele leito de hospital. Envolveu o pai com o um abraço e, em meio a um beijo suave disse: "vim te buscar".

Vestiu o terno de linho branco, a muito esquecido, e de mãos dadas partiu com a filha.

Do lado de fora um alvoroço: polícia, bombeiros, muita correria e gente desesperada. Não se sabe como, a favela havia incendiado e já se contavam dezenas de mortos. Ele nem ligava. Já se passavam das dez... Estava no paraíso!

Fonte: <http://www.overmundo.com.br/>

Nesta publicação o C2 postou as seguintes tags: "miseria", "dor", "sofrimento", "tristeza", "solidao", "melancolia", "vicios", "morte", "abandono", "ficcao".

As categorias propostas por Pejtersen (1978 apud MOREIRA, 2006) em sua já mencionada grade de metodologia de análise de assunto para a indexação dos textos literários foram preenchidas da seguinte forma:

Quadro 11 - Distribuição das tags da publicação 4 classificadas na grade de classificação proposta por Pejtersen (1978 apud MOREIRA, 2006)

USUÁRIO:C2				
TÍTULO: Paraíso				
CATEGORIAS DE PEJTERSEN				
Assunto O QUE ?	Personagens QUEM?	Estrutura/ Cenário ONDE?	Intenção do Autor COMO?	Acessibilidade QUAL?
Morte			miseria dor sofrimento tristeza solidao, melancolia vicios abandono	ficção
TAGS: miseria, dor, sofrimento, tristeza, solidao, melancolia, vicios, morte, abandono, ficcao				

Fonte: Moreira (2006) com adaptação da autora

Análise de publicação 5

O AMOR ENLOQUECE

Viver enobrece !
Sem amar, a Alma enegrece !
Olhos infantis, apenas envaidecem !
Furores juvenis, intumescem...
Emoções que se aquecem !
A Fome apetece !

As pessoas crescem, de atenções carecem !
Muito frequentemente se enfurecem !
De vez em sempre se enternecem !
Contra os problemas se guarnecem !
Invariavelmente se aborrecem !
Os imaturos amadurecem !
As Químicas favorecem !
Os cabelos embranquecem !
A Pele enrigesce !
A cútis empalidece !
O aspecto rejuvenesce !
A gente não esmorece !
A Afeição é como se nunca houvesse...
O amor se desvanece !
Os amigos comparecem !
Perguntas sem respostas se oferecem !
Os Sentimentos se embrutecem !
Mortes que nos endurecem !
Um gosto de Nada se fornece !
Ah, se eu ainda pudesse... mas as forças se enfraquecem, o duro amolece, o Instinto desobedece, as ofertas desaparecem, o ciúme empurta !
A garota de mim se esquece !
O jogo fenece!
Um Silêncio nos ensurdece !
O Olhar que ensandece !
O Adeus aparece !
O corpo ? ah, como esse padece... e o que permanece ?
Amar enlouquece !

Fonte: <http://www.overmundo.com.br/>

Nesta publicação o C3 postou as seguintes tags: "amor", "loucura", "pessoa", "desilusao", "rima", "poesia".

As categorias propostas por Pejtersen (1978 apud MOREIRA, 2006) em sua já mencionada grade de metodologia de análise de assunto para a indexação dos textos literários foram preenchidas da seguinte forma:

Quadro 12 - Distribuição das tags da publicação 5 classificadas na grade de classificação proposta por Pejtersen (1978 *apud* MOREIRA, 2006)

USUÁRIO:C3				
TÍTULO: O amor enloquece				
CATEGORIAS DE PEJTERSEN				
Assunto O QUE ?	Personagens QUEM?	Estrutura/ Cenário ONDE?	Intenção do Autor COMO?	Acessibilidade QUAL?
loucura amor	Pessoa		desilusao rima	poesia
TAGS: amor, loucura, pessoa, desilusao, rima, poesia				

Fonte: Moreira (2006) com adaptação da autora

Análise de publicação 6

FESTA

Tá rolando aqui dentro uma festa que se manifesta
no centro do meu coração!
Como na primeira vez, a gente se testa,
ou algo fala ou nada diz!
De você eu gosto,
só que não me encosto pra te beijar!
Eu te dou um tempo,
mas estou atento nesse tal gostar!
E nessa festa te amo
e sei te dar o devido valor!
Nada se contesta,
confie nos meus gestos e naquilo que atesto!
Vê se não protesta, desenruga a testa,
você não vai se enganar!
Não seja indigesta, menina modesta,
deixa eu te embriagar...
Come, a gente não presta,
fujamos pra floresta num intervalo trivial!
E pela mesma frestadeira não funesta ocasião,
a gente empresta um pouco da nossa Razão...
Continua a festa, intensa,
louca & honesta, interessante ao extremo!

Tá rolando aqui dentro uma festa
e eu não sei como te convidar!

Fonte: <http://www.overmundo.com.br/>

Nesta publicação o C3 postou as seguintes *tags*: declaracao-de-amor, relacionamento, namoro, enamorado, casal, beijo, poesia.

As categorias propostas por Pejtersen (1978 apud MOREIRA, 2006) em sua já mencionada grade de metodologia de análise de assunto para a indexação dos textos literários foram preenchidas da seguinte forma:

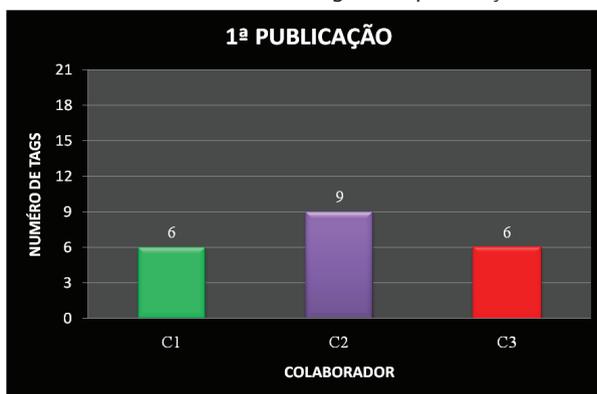
Quadro 13 - Distribuição das tags da publicação 6 classificadas na grade de classificação proposta por Pejtersen (1978 apud MOREIRA, 2006)

USUÁRIO: C3				
TÍTULO: Festa				
CATEGORIAS DE PEJTERSEN				
Assunto O QUE ?	Personagens QUEM?	Estrutura/Cenário ONDE?	Intenção do Autor COMO?	Acessibilidade QUAL?
declaracao-de-amor	Casal		relacionamento namoro enamorado	poesia
TAGS: declaracao-de- amor, relacionamento, namoro, enamorado, casal, beijo, poesia				

Fonte: Moreira (2006) com adaptação da autora

A partir da análise das 1ª e das últimas publicações podemos estabelecer gráficos comparativos do número de *tags* postadas por cada colaborador, como vemos na gráfico 1, onde o C1 postou 6 *tags*, o C2 postou 9 *tags* e o C3 postou 6 *tags*, resultando em um número de 21 *tags* na 1ª publicação.

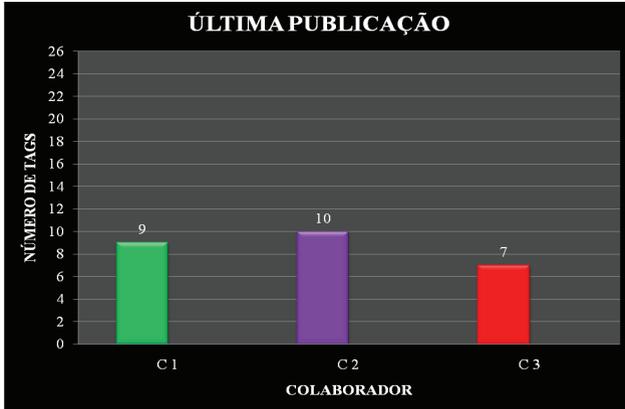
Gráfico 1 - Número de tags da 1ª publicação



Fonte: Elaborado pela autora

Já no gráfico 2 pode-se constatar um número maior de *tags* postada pelos colaboradores, cujo o gráfico mostra que o C1 postou 9 *tags*, o C2 postou 10 *tags* e o C3 postou 7 *tags*, resultando um número de 26 *tags* postadas na última publicação.

Gráfico 2 - Número de tags da última publicação

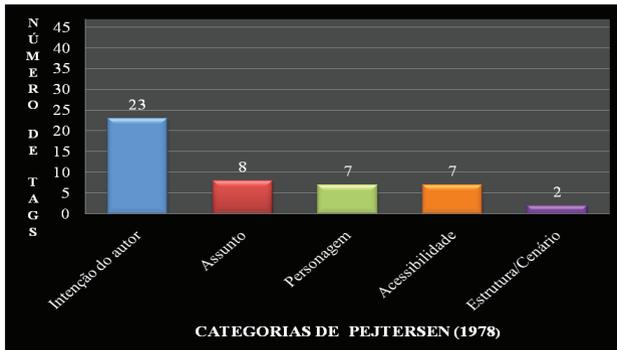


Fonte: Elaborado pela autora

Esta comparação nos mostra que o usuário/colaborador evoluiu de forma mais segura na postagem de *tags* aos seus conteúdos, haja vista que o mesmo seguiu as recomendações sugeridas pelo sistema do Overmundo, descartando assim, a geração de falhas correntes em sistemas de indexação, como a polissemia, sinonímia ente outros.

O gráfico 3 nos mostra como as 47 *tags* analisadas foram distribuídas segundo as categorias proposta por Pejtersen (1978 apud MOREIRA, 2006).

Gráfico 3 - Distribuição das tags classificada entre as categorias propostas por Pejtersen (1978 apud MOREIRA, 2006)



Fonte: Elaborado pela autora

De acordo com o gráfico acima a maior parte das tags corresponde à categoria “intenção do autor” com 23 etiquetas (49%); a categoria “assunto” ocupa o 2º lugar com 8 etiquetas (17%); as categorias “personagem” (15%) e “acessibilidade”(15%) contam com o mesmo número etiquetas, possuindo 7 etiquetas a categoria “estrutura/cenário” ocupa a última posição com 2 etiquetas (4%).

As tags são bem distribuídas entre a maior parte das categorias, com exceção da categoria “intenção do autor”, que recebeu um número maior de tags.

Os usuários/colaboradores atribuíram um número alto de *tags* (23), que apontaram para à categoria “intenção do autor”, constatamos a partir daí que por se tratar de algo tão subjetivo, requerendo interpretação pessoal de cada indexador/etiquetador, poderia ser utilizada para expressar o que Beghtol (1986 apud MOREIRA 2006, p. 102) chama de significados do texto (meanings). O termo foi utilizado para os significados representativos de um contexto de análise, onde o indexador/etiquetador identificou e selecionou o que é significativo para sua publicação no caso da pesquisa em questão. Tal afirmativa fica clara na resposta do C (1) quando escreve: “As *tags* normalmente chegam com a poesia, faço referência, sempre, ao que escrevi. Entendo que fica mais coeso [...]”.

Portanto, fica claro que as *tags* postadas nas publicações analisadas foram motivadas por critérios conotativos (intenção do autor) que representam elementos interpretados pelo próprio usuário/colaborador.

A categoria “assunto” que ocupou o 2º lugar como critério de motivação, para os usuários/colaboradores, realmente é a categoria que abriga a palavra mágica, a atinência. Foi possível observar isso durante sua utilização quando o C 3 afirma que “Meu maior medo (creio que já deve ter ocorrido vez ou outra), é dar o mesmo assunto a textos diferentes, rssssss!”

Após a categoria assunto aparecem as categorias “personagem”, “acessibilidade” e “estrutura/cenário” com um baixo número de termos atribuídos pelos usuários/colaboradores. Como explica o C 2 “[...] faço uso daquelas que facilitam o acesso das pessoas aos trabalhos publicados” e ainda “[...] quando as uso faço conforme a natureza de meus trabalhos[...]”. Observa-se, portanto, que por estas categorias representarem elementos reais, na forma denotativa, e ainda serem identificados como dados objetivos de um texto, podem não ser relevantes para os usuários/colaboradores do Overmundo no momento de *tagear* suas publicações.

REFERÊNCIAS

ALBUQUERQUE, A.; MESQUITA, D.; COSTA, L. Folksonomia: uma nova modalidade de indexação e recuperação da informação na web. In: Encontro Nacional de Estudantes de Biblioteconomia, Documentação, Gestão e Ciência da Informação, n. 33, 2010, João Pessoa. **Anais...** João Pessoa: ENEBD, 2010.

AQUINO, M. C. A potencialização da memória coletiva através do hipertexto na Web 2.0. In: Congresso Brasileiro de Ciências da Comunicação, 30., 2007, Santos. **Anais...** Santos: Sociedade Brasileira de Estudos Interdisciplinares da Comunicação, 2007.

BEGHTOL, C. A. Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. **Journal of Documentation**, London, v. 42, n. 2, p. 84-113, jun. 1986.

BELLEI, S. L. P. **O livro, a literatura eo computador**. São Paulo: EDUC; Florianópolis: UFSC, 2002.

BRANDT, Mariana Baptista. **Etiquetagem e folksonomia**: uma análise sobre a óptica dos processos de organização e recuperação da informação na web. 144f. 2009. Dissertação (Mestrado em Ciência da Informação) - Universidade de Brasília, Brasília, 2009.

BRANDT, Mariana; MEDEIROS, Marisa Bräscher Basílio. Folksonomia: esquema de representação do conhecimento. **Transinformação**, Campinas, v. 22, n. 2, p. 111-121, maio/ago. 2010.

CALDAS, W. F; MOREIRA, M. P. Folksonomia e classificação de etiquetas: estudos de caso. In: Encontro Nacional de Pesquisa em Ciência da Informação, 10., 2009, João Pessoa. **Anais...** João Pessoa: ENANCIB, 2009.

CARLAN, Eliana. **Sistemas de Organização do Conhecimento**: uma reflexão no contexto da Ciência da Informação. 195f. 2010. Dissertação (Mestrado em Ciência da Informação) - Universidade de Brasília, Brasília, 2010.

CARVALHO, Lidiane dos Santos; LUCAS, Elaine R. de Oliveira; GONCALVES, Lucas Henrique. Organização da informação para recuperação em redes de produção e colaboração na web. **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis, v. 15, n. 1, p. 71-86, jan./jun. 2010.

CATARINO, Maria Elizabete; BAPTISTA, Ana Alice. Folksonomia: um novo conceito para a organização dos recursos digitais na Web. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 2, n. 2, 2007.

CHRISTIANS, C. G. A ética e a política na pesquisa qualitativa. In: DENZIN, N. K.; LINCOLN, Y. S. (ed.). **O planejamento da pesquisa qualitativa: teorias e abordagens**. Porto Alegre: Artmed, 2006. p.141-162.

COSTELLA, Antonio F. **Comunicação: do grito ao satélite**. 4. ed. Campos do Jordão: Mangueira, 2001.

DODEBEI, V. L. D. **Tesouro: linguagem de representação da documentária**. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002.

FEITOSA, Ailton. **Organização da informação na web: das tags à web semântica**, Brasília: Thesaurus, 2006.

FIGUEIREDO, Nélia Maria Almeida de (Org.). **Método e metodologia na pesquisa científica**. 3. ed. São Caetano do Sul, SP: Yendis Editora, 2008.

GOLDER, S.; HUBERMAN, B. A. Usage patterns of collaborative tagging Systems. **Journal of Information Science**, v. 32, n. 2, p. 198-208, 2006.

GOUVÉA, Cleber; LOH, Stanley. Folksonomias: identificação de padrões na seleção de tags para descrever conteúdos. **RESI - Revista Eletrônica de Sistemas de Informação**, v. 11, n. 2, 2007.

GUIMARÃES, Rafael Costa. **Uso da folksonomia e da etiquetagem na indexação de imagens**. 88f. 2012. Monografia (Graduação em Biblioteconomia) - Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2012.

HAYKIN, Simon; MOHER, Michael. **Sistemas modernos de comunicações wireless**. Porto Alegre: Bookman, 2008.

JOHNSON, Steven. **Cultura da interface: como o computador transforma nossa maneira de criar e comunicar**. Rio de Janeiro: Jorge Zahar Editor, 2001.

KOZINETTS, Robert V. The field behind the screen: using netnography for marketing research in online communities. **Journal of Marketing Research**, Chicago, v. 39, n. 1, fev. 2002.

LACERDA, J. A. C. de; VALENTE, P. G. A emergência em sistemas baseados em folksonomias. In: **Estudos em Jornalismo e Mídia**. Ano IV. n. 2- 2º semestre de 2007.

LANCASTER, F. W. Indexação e resumos: teoria e prática. Brasília: Briquet de Lemos Livros, 1993. LUGER, G. F. **Inteligência artificial: estruturas e estratégias para a solução de problemas complexos**. 4 ed. Porto Alegre: Bookman, 2004.

MICHAELIS ONLINE. **Significado de literatura**. 2017. Disponível em: <https://michaelis.uol.com.br/>.

MINAYO, M. C. S. Ciência, técnica e arte: o desafio da pesquisa social. In:2 (Org.). **Pesquisa social: teoria, método e criatividade**. 18. ed, Petrópolis: Vozes, 1994.

MOREIRA, Margareth Egídia. **Análise de assunto da literatura ficcional infantil: categorias para ler o que você tem**. 103f. 2006. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

MOTA, Francisca Rosaline Leite; OLIVEIRA, Marlene de. Formação e atuação profissional. In: OLIVEIRA, Marlene de (Coord.). **Ciência da Informação e Biblioteconomia: novos conteúdos e espaços de atuação**. Belo Horizonte: Ed. UFMG, 2005. p. 97-110.

MUELLER, Suzana Pinheiro Machado. **Métodos para a pesquisa em Ciência da Informação**. Brasília: Thesaurus, 2007.

NUNES JÚNIOR, Marco. **Nuvem de tags**. 2009. Disponível em: <http://Acessoblogs.ambientelivre.com.br/marcio/nuvens-de-tags-online>. em: 30 abr. 2021

NASCIMENTO, G. F. C. de L. **Folksonomia como estratégia de indexação dos bibliotecários no Del.icio.us**. João Pessoa: PPGCI, 2008.

NAVES, Madalena Martins Lopes. Análise de assunto: concepções. **Revista de Biblioteconomia de Brasília**, v. 20, n. 2,p. 215-226, jul./ dez. 1996.

NIELSEN, Hans Jørn. The nature of fiction and its significance for classification and indexing. **Information Services & Use**, v. 17, n. 2/3, p. 171-181, 1997. (Special Issue: Eletronic access to fiction).

OLIVEIRA, Hellen Carmo de; CARVALHO, Cedric Luiz de. **Gestão e representação do conhecimento**. Goiânia: UFG, 2008.

OVERMUNDO. **Página do usuário**. 2013. Disponível em: <http://www.overmundo.com.br/registro/login.php>. Acesso em: 13 mar. 2021.

PASSARELLI, B. O bibliotecário 2.0 e a emergência de novos perfis profissionais. **DataGramZero**, Rio de Janeiro, v. 10, n. 6, novdez., 2009.

PEJTERSEN, Annelise Mark. Fiction and library classification. **Scandinavian Public Library Quartely**, n. 1, p. 5-12, 1978.

PEJTERSEN, Annelise Mark. Fiction and library classification. **Scandinavian Public Library Quarterly**, n. 1,p. 5-12, 1978.

PEREIRA, Henrique C. **Folksonomia e a maneira com que nós colocamos ordem nas coisas**. Disponível em: <http://www.revolucao.etcbr/archives/folksonomia-e-a-maneira-com-que-ngs-colocamos-ordem-nas-coisas/>. Acesso em: 02 jun. 2021.

PIEIDADE, M. A. R. **Introdução a teoria da classificação**. 2 ed. Rio de Janeiro: Interciência, 1983.

PRIMO, Alex. O aspecto relacional das interações na web 2.0. In: INTERCOM: Congresso Brasileiro de Ciências da Comunicação, 29., 2006, Brasília. **Anais...** Brasília: INTERCOM, 2006.

QUINTARELLI, E. Folksonomies: power to the people. In: ISKO, 2005, Milano. **Anais...** Milano: ISKO, 2005.

SOUSA, Denner Santiago de; RUFINO, Sandra. Tecnologias sociais: panorama da Universidade Federal do Rio Grande do Norte. **Revista tecnologia e sociedade**, Universidade Federal do Rio Grande do Norte: Natal, 2010

SILVA, Fabiano Couto Corrêa da; BLATTMANN, Ursula. A colaboração a interação na web 2.0. In: Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação, 22., 2007, Brasília. **Anais...** Brasília: CBBB, 2007.

SILVA, J. F. M. O impacto tecnológico no exercício profissional. In: VALENTIM, M. L. P. (Org.). **Atuação profissional na área de informação**. São Paulo: Polis, 2004. p. 83-96.

STAKOVIK, Henrique Moura . **Implantação de um mecanismo de enriquecimento do perfil do usuário e recomendação de trabalhos científicos para o konnen** 201174 F. Trabalho de Conclusão de Curso (Graduação de Sistema de informação) - Centro Universitário Luterano de Palmas, Palmas, 2011.

SVENONIUS, Elaine. **A base intelectual da organização da informação**. Cambridge: The Mit Press, 2000.

VALENTIM, Marta Lígia Pomim. Atuação e perspectivas profissionais do profissional da informação. In: **O profissional da informação: formação, perfil e atuação profissional**. São Paulo: Polis, 2000. p. 135-152. (Coleção palavra-chave, 11).

VIDICH, A.; LYMAN, S. Métodos qualitativos: sua história na Sociologia e na Antropologia. In: DENZIN, N. K.; LINCOLN, Y. S. (Orgs.). **O planejamento da pesquisa qualitativa: teorias e abordagens**. 2. ed. Porto Alegre: Artmed, 2006.

VIERA, Angel Freddy Godoy; GARRIDO, Isadora dos Santos. Folksonomia como uma estratégia para recuperação colaborativa da informação. **Data GramZero- Revista de Ciência da Informação**, v. 12, n. 2, abr. 2011.

WAL, T. W. **Folksonomy**. [S.l.]: [S. n.], 2007. Disponível em: <http://www.vanderwal.net/folksonomy.html>. Acesso em: 22 jul. 2021.

WAL, T. W. **Folksonomy**. 2005. Disponível em: <http://www.vanderwal.net/essays/051130/folksonomy.pdf>. Acesso em: 22 abr. 2021.

WARD, K. J. The cyber-ethnographic (re)construction of two feminist online communities. **Sociological Research Online**, v. 4, n. 1, 1999.



UTILIZAÇÃO DE ONTOLOGIAS NA ORGANIZAÇÃO DA INFORMAÇÃO NAS BIBLIOTECAS UNIVERSITÁRIAS FEDERAIS DAS REGIÕES NORTE E NORDESTE DO BRASIL

Sandra Gleice da Silva e Silva

INTRODUÇÃO

O contexto atual caracteriza-se pelo crescimento desenfreado da informação e por um crescente volume da produção científica, juntamente a isto, cresce a valorização da informação. Neste cenário a internet constitui-se como uma ferramenta que tem revolucionado todas as áreas do conhecimento pela facilidade de disponibilização de informação e o volume de recursos que cresce a cada dia.

Após o advento do serviço *World Wide Web* na internet houve um grande crescimento no volume de informações disponíveis, proporcionando um maior número de usuários ligados à rede. No entanto, juntamente a esse volume de informações surgiram alguns problemas relacionados ao acesso e recuperação, surgindo então, a necessidade de se organizar o conteúdo disponibilizado, de forma que a informação fosse filtrada e chegasse ao usuário de forma mais precisa.

Essa grande disponibilidade de informação gera alguns problemas de busca e recuperação, como a falta de precisão. Surge, então, a necessidade do refinamento de buscas e a Web Semântica tem sido indicada como solução para representação dos recursos informacionais na *web*, face às dificuldades de encontrar informações adequadas às necessidades dos usuários.

São inúmeras as dificuldades encontradas pelos usuários no momento da recuperação da informação, que variam desde a falta de conhecimento por parte dos mesmos a ineficácia dos mecanismos de busca. O fato é que embora a *web* seja um enorme repositório de informações, estas se encontram de forma desestruturada e muitas vezes sem nenhum sentido, surgindo a necessidade de instrumentos que permitam a padronização de metodologias de estruturação de páginas e *sites*, como a Web Semântica, que segundo Feitosa (2006), é um método de gravação de significados em páginas da *web* que se unem por meio de um esquema que adiciona interpretação lógica aos documentos.

Embora a internet utilize uma linguagem de conhecimento comum, o *Hypertext Markup Language* (HTML) que permite que indivíduos produzam livremente seus conteúdos, de forma facilitada e independente de sua localização geográfica, essa produção de conteúdos acabou gerando um crescimento desordenado da *web*. Dessa forma ficou mais difícil a recuperação de informação útil, que se deve tanto às falhas dos navegadores que segundo Mamede (2006), capturam apenas a parte léxica, surgindo então, a necessidade de considerar a questão semântica, cuja captura baseia-se no conteúdo.

Para a modelagem da informação e sua melhor organização surgiu a Web Semântica cuja finalidade é acrescentar significado aos documentos disponíveis na rede e “[...] Nesse contexto, a Ontologia ocupa função preponderante como forma de representação do conhecimento, de modo a proporcionar seu compartilhamento, reutilização e manutenção”. (MAMEDE, 2006, p.15).

A intenção inicial da Web Semântica é acrescentar semântica ao conteúdo da *web*, a fim de determinar o sentido de um texto no contexto de um determinado documento, facilitando e melhorando a recuperação da informação relevante. Tendo as Ontologias como um fator chave nesse processo, visto que possibilitam um padrão para o compartilhamento da informação.

A inserção dos bibliotecários nesse novo cenário tem exigido nova postura, novos saberes, visto que é denominado um profissional da informação, e para tal, necessita conhecer os novos instrumentos de representação, a fim de organizar a informação para atender, de forma satisfatória aos seus usuários.

Surge, então a necessidade de estudar o processo de representação e organização da informação por meio de Ontologias nas Bibliotecas Universitárias Federais das regiões Norte e Nordeste do Brasil, na perspecti-

va de conhecer como os bibliotecários se posicionam diante desse novo cenário informacional, seu grau de familiaridade e uso desse instrumento na representação e organização de seus Repositórios Institucionais e como estas instituições estão contribuindo para o processo de organização da informação na *web*.

Para fundamentar o estudo recorreu-se a autores como Silva e Sales (2011), que apontam as Ontologias como um forte instrumento semântico para dar suporte à interoperabilidade entre sistemas de informação. Anjos (2008), que trata sobre a Teoria da Classificação e a organização do conhecimento em Biblioteconomia e Ciência da Informação. Piedade (1983) e Barbosa (1969), por serem os clássicos na área, e seus conceitos sobre as classificações bibliográficas e a organização do conhecimento nas estantes.

Espera-se que este estudo possa gerar informações úteis para aprimorar os conhecimentos dos profissionais bibliotecários, bem como uma base de dados de Ontologias disponíveis e produzidas na área da Ciência da Informação.

ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

O contexto informacional atual apresenta grandes desafios, diante da crescente explosão informacional gerada pela *web*, que possibilitou um novo meio de interação e comunicação. Segundo Dziekaniak (2010), no novo cenário projetado acerca da organização da informação cabe salientar a presença de duas revoluções:

[...] a documentária, processo pelo qual passou e passa a comunidade científica mundial, através da expressividade crescente da publicação de documentos, principalmente através dos periódicos e, principalmente a revolução tecnológica, focalizada no âmbito das tecnologias da informação e comunicação fornecendo aparato tecnológico para a produção e disseminação da informação.

As tecnologias da informação e comunicação (TIC) têm alterado sobremaneira o *status quo* tanto da produção quanto do acesso à informação. Diante disso, surge a necessidade de ferramentas capazes de extrair a informação precisa entre inúmeros dados relevantes.

O modelo de representação nas bibliotecas tradicionais necessita passar por mudanças, pois o mesmo deixa a desejar em relação à internet, devido ao crescimento exponencial da produção de conhecimento. A biblioteca como possuidora de um público de usuários heterogêneo, deve estar munida de fazeres e saberes que atendam às necessidades informacionais.

A informação disponibilizada na *web*, embora apresente uma face com usabilidade, não tem atendido de forma satisfatória os seus usuários, devido à mesma ser indexada por palavras-chave que não permitem que se identifique o contexto da informação procurada. Souza e Alvarenga (2004), afirmam que a *web* “[...] foi implementada de forma descentralizada e quase anárquica, cresceu de maneira exponencial e se apresenta atualmente como um imenso repositório de documentos que deixa muito a desejar quando se trata de recuperação de conteúdo relevante”.

Diante desse caos informacional surge a Web Semântica caracterizada como uma extensão da *web* atual, objetivando aprimorar o serviço de busca e recuperação na *web*. Para Pickler (2007), “[...] a Web Semântica visa, justamente, melhorar a satisfação do usuário no momento da busca, retornando-lhe as informações adequadas à suas necessidades”.

As Ontologias fazem parte dos elementos que compõem a Web Semântica, designando a faceta semântica da representação dos seres. A Ontologia seria então, como uma especificação de uma conceituação e segundo Oliveira (2002):

A Ontologia, na web semântica estabelece uma ligação terminológica entre membros de uma comunidade podendo ser estes membros, agentes humanos ou máquinas. No jargão dos pesquisadores em inteligência artificial, uma ontologia é um documento ou arquivo que define formalmente a relação entre termos.

As Ontologias são a língua franca da Web Semântica, possibilitando a elaboração de uma rede de reconhecimento humano a fim de melhorar o nível de serviços na *web*, garantindo uma comunicação livre de ambiguidades. Segundo Silva e Sales (2011), “[...] constituem um meio poderoso de inter-relacionar dados em sistemas de informação [...] são elaboradas, principalmente, visando à estruturação de bases de conhecimento ou para serem utilizadas como ferramentas semânticas no suporte à interoperabilidade entre sistemas de informação.

Silva e Sales (2011), afirmam que Ontologia é um conjunto de conceitos padronizados onde termos e definições devem ser aceitos por uma comunidade no âmbito de um domínio [...] permitem a elaboração de estruturas tanto terminológicas quanto taxionômicas. Os bibliotecários devem lançar vistas ao estudo de tecnologias e das novas formas de representação e organização sob a óptica da Web Semântica e o uso de Ontologias no gerenciamento de dados. Além de se conectar com os novos promotores de informação como os Repositórios Institucionais e os Arquivos Abertos, revendo seu papel e pensando no ambiente eletrônico/digital.

Segundo Sales (2011), os domínios são representados em dois momentos: na organização da informação e na organização do conhecimento. A organização da informação no contexto da Biblioteconomia consiste nas atividades de classificação, catalogação e indexação de documentos e a organização do conhecimento é realizada por meio dos esquemas de classificações, os tesouros e as Ontologias.

No âmbito da organização da informação a catalogação utiliza os códigos de catalogação, cuja base teórica reside nos princípios traçados por C.A Cutter. A classificação adota os sistemas de classificação decimal com base teórica em M. Dewey. E a indexação por sua vez, utiliza-se de tesouros, taxonomias, Ontologias e vocabulários controlados.

A organização do conhecimento pressupõe a construção de modelos de representação do mundo. Tem por objeto o conjunto das ideias, convertido em conceitos ou aglomerado destes, os quais devem ser analisados para a verificação de seus vínculos a um domínio de conhecimento. A intenção é confeccionar protótipos de mundo via representação dos conhecimentos de uma coletividade. O produto subsequente é a representação do conhecimento em um dado campo, e não somente a representação da informação. (SALES, p. 72, 2011).

Sales (2011, p.62), afirma que “[...] domínios são reflexos discursivos [...] passíveis de tratamento no âmbito da Ciência da Informação mediante sua materialização em forma de texto e sua institucionalização e socialização em forma de documento [...]”. E no contexto da Biblioteconomia esses domínios têm a informação organizada por meio dos processos de catalogação, classificação e indexação e o conhecimento por meio dos sistemas de classificação, cabeçalhos de assuntos, taxonomias, tesouros e Ontologias.

A organização e representação do conhecimento desses domínios materializam a informação, conferindo-lhe característica de materialidade institucional e social, visando à codificação e decodificação tanto por profissionais, quanto por usuários da informação.

No contexto da Ciência da Informação é necessário organizar para representar e a organização da informação está ligada a análise e tradução dos assuntos dos documentos para uma linguagem que melhor se adeque ao sistema de informação. O mundo de informações cresceu rapidamente com o advento da internet e o bibliotecário já não é mais o único mediador do acesso, surge então, a necessidade de se adequar às novas formas de organização da informação e do conhecimento, munindo-se das ferramentas adequadas à nova realidade ambiental.

A organização do conhecimento é uma forma de representação do mundo e a organização da informação está ligada às tarefas de análise e tradução dos assuntos de um determinado documento para uma linguagem adequada a um sistema de informação. Cervantes (2006, p. 9), afirma que:

Um sistema de informação pode ser uma biblioteca pública ou especializada; um centro de documentação de uma empresa; um arquivo; um museu; ou um banco de dados. Independente de sua denominação, um sistema de informação tem como principal função coletar, tratar e disseminar a informação produzida ou incorporada pela sociedade na qual está inserido, assegurando o acesso à cultura por parte de seus membros e possibilitando sua continuidade.

O cenário atual apresenta inúmeros sistemas de informação que possibilitam aos usuários o acesso facilitado, no entanto, ainda limitado a informações sem precisão e coerência, cabendo ao bibliotecário buscar capacitação contínua a fim de estar munido das habilidades necessárias ao exercício de suas atividades enquanto organizador e disseminador da informação e do conhecimento.

Aspectos históricos da organização do conhecimento científico

O ato de classificar esteve presente desde a Antiguidade, trata-se de um processo mental, por meio do qual se faz o agrupamento por semelhanças e diferenças. A classificação bibliográfica atual sofreu influências tanto dos sistemas de classificação antigos, quanto dos sistemas de classificação filosófica (Quadro 1).

Segundo Barbosa (1969, p. 43), “Classificações filosóficas são as puramente teóricas, constituindo agrupamentos dos conhecimentos humanos segundo o ponto de vista de seus idealizadores.” Platão foi o primeiro a agrupar os conhecimentos humanos segundo bases filosóficas, dividindo-os em Física, Ética e Lógica.

O filósofo grego Aristóteles (384 – 322 a. C.), discípulo de Platão, propôs a divisão tritônica do conhecimento em três filosofias ou ciências: teóricas (filosofia e matemática), práticas (ciências exatas e sociais) e produtivas (arte e literatura). Equivalentes às faculdades exercidas pelo homem: pensar, agir e produzir.

Anjos (2008, p.45), afirma que “[...] a classificação tritônica de Aristóteles (Ciências Teóricas, Práticas e Produtivas) juntamente com a classificação tritônica de Platão (Física, Ética e Lógica), [...] influenciarão a classificação dos conhecimentos que se formulará na Idade Média”.

Outra grande contribuição na organização do conhecimento foi a classificação de Porfírio. Souza (2010), afirma que Porfírio (ano 305 d.C.) dividiu o conhecimento, embasado no princípio dedutivo da filosofia. Seu sistema, conhecido como Árvore de Porfírio é o primeiro exemplo de uma classificação binária, partindo de termos de grande extensão e pouca intenção para termos de pouca extensão e grande intensão, ou seja, do geral para o específico.

Francis Bacon (1561 – 1626) idealizou uma classificação das ciências baseada nas faculdades humanas da Memória, Imaginação e Razão, originando respectivamente a História, Poesia e Filosofia. Sua classificação é a mais significativa dentre os precursores da ciência moderna, servindo de fundamento para a construção de vários instrumentos de organização do conhecimento como a *Encyclopédie* de Diderot e d'Alembert e de várias classificações bibliográficas como a de Harris, de Dewey, Classificação da Biblioteca do Congresso dos Estados Unidos e a de Otlet.

No final do século XVII Leibniz (1646 – 1716), encarregado da biblioteca do duque de Wolfenbüttel, alinhou o conhecimento filosófico com o bibliotecário na organização desta biblioteca e produziu um “Plano para organizar uma biblioteca” que “[...] dividia o conhecimento em nove partes, três das quais correspondiam às tradicionais faculdades superiores de teologia, direito e medicina, acrescidas de filosofia, matemática, física, filologia, história e assuntos diversos.” (PETER, 2003, p.98).

Os filósofos iluministas Diderot (1713 – 1784) e d'Alembert (1717 – 1783) embasados nos princípios da classificação de Bacon (Memória, Imaginação e Razão) mudam a ordem para Memória, Razão e Imaginação e sistematizam a “Enciclopédia” que reunia o conhecimento e o pensamento filosófico da época, além de ter sido a primeira enciclopédia alfabética das ciências, cuja preocupação estava centrada na exposição sistemática das ciências e não em sua estrutura hierárquica.

Ao longo da história foram criados inúmeros esquemas para reformular a classificação do conhecimento, os esquemas apresentados pelos filósofos trouxeram soluções pragmáticas que não resolviam todos os problemas relacionados à organização do conhecimento, surgindo a necessidade da criação de sistemas de classificação mais flexíveis, visando uma melhor organização no gerenciamento da informação e sua disseminação.

Quadro 1- Classificações Filosóficas

PERÍODO	FILÓSOFO	DIVISÃO DAS CIÊNCIAS
Séc. IV a. C.	Platão	Física, Ética, Lógica
Séc. IV a. C.	Aristóteles	Filosofia Teórica, Filosofia Prática, Filosofia Poética
Séc. VI	Cassiodoro	Ciências Sermoniais, Ciências Reais
Séc. XVII	F. Bacon	Ciências da Memória, Ciências da Razão, Ciências da Imaginação
Séc. XVII	Locke	Ciências Reais, Ciências Ideais
Séc. XVIII	Diderot e d'Alambert	História, Filosofia, Poesia
Séc. XIX	Hegel	Ciências do Absoluto, Ciências da Natureza, Ciências do Espírito
Séc. XIX	Ampère	Ciências Cósmicas, Ciências Noológicas
Séc. XIX	Comte	Ciências Abstratas Fundamentais, Ciências Concretas Derivadas
Séc. XIX	Spencer	Ciências Abstratas, Ciências Abstrato-concretas, Ciências Concretas
Séc. XX	Wundt	Ciências Formais, Ciências Reais: Natureza e Espírito

Fonte: Própria da pesquisa

As classificações filosóficas além de fazer a divisão das ciências, bem como a organização dos saberes em suas classes e categorias permitiam relacioná-los entre si e especificar suas diferenças, influenciando várias classificações bibliográficas. Assim como as classificações de Platão e Aristóteles influenciaram as classificações da Idade Média, Francis Bacon e seu princípio usado na divisão do conhecimento foram um marco que continua sendo usado até hoje, e uma das maiores influências nas classificações bibliográficas.

Segundo Anjos (2008), a prática de dividir e organizar documentos remonta desde os anos de 260 e 240 a. C. quando o poeta e sábio grego Calímaco, chefe da Biblioteca de Alexandria e também considerado o primeiro bibliotecário da história, publicou um catálogo (Panakes) no qual dividia os livros por tipo de escritor. O esquema classificatório de Calímaco organizou cerca de 500 mil volumes da Biblioteca dos Ptolomeus, que era organizada por dois índices, um de autor e outro de título.

Cutter (1837 – 1903), bibliotecário americano desenvolveu a *Expansive classification* (EC) que serviu de base para a classificação inicial da Biblioteca do Congresso. “[...] Sua notação é alfabética, onde as classes principais são representadas por letras maiúsculas e suas subdivisões por letras minúsculas. Os números são utilizados para as subdivisões de forma, história e geografia.” (SOUZA, 2010, p.18).

Julius Kaiser (1868 – 1927) bibliotecário norte-americano em sua obra *Systematic Indexing* afirma que a indexação permite o agrupamento de assuntos semelhantes, apresentando as bases para o tratamento temático facetado, sendo, portanto, o precursor da classificação facetada de Ranganathan. Para Kaiser, todo assunto é composto de duas categorias básicas: o concreto e o processo, e a combinação destes é denominada enunciado. Estas duas categorias influenciaram Ranganathan nas categorias personalidade e energia, respectivamente.

A base do esquema da Classificação Decimal de Dewey (CDD) foi influenciada por Harris que por sua vez foi influenciado por Bacon, possui uma notação bem fácil, baseada numa sequência decimal e utilizando letras. A CDD dividiu o conhecimento humano em dez classes que vão se subdividindo, sempre decimalmente, possuindo divisões de forma, língua, literatura, geográficas e de tempo. Amplamente utilizada em todo o mundo, especialmente em bibliotecas públicas.

A Classificação Decimal Universal (CDU) foi desenvolvida por Paul Otlet e Henri La Fontaine, baseou-se na CDD (Quadro 2) e expandida a tabelas auxiliares, possuindo uma notação de cada classe feita por apenas um algarismo e cada classe se subdivide em classes de dois algarismos, além de apresentar uma estrutura com tabelas principais, tabelas auxiliares e índice alfabético. Geralmente é utilizada em bibliotecas especializadas. Dentre os esquemas de classificação tradicional a CDU é a única que não se limita apenas à organização física dos documentos na estante, foi planejada para reunir documentos sobre o mesmo assunto (princípio monográfico de Otlet).

A *Bibliographic classification* (BC) esboçada por Henry Evelyn Bliss reúne o conhecimento humano em quatro grupos: Filosofia – Ciência – História – Tecnologia e Artes, que se subdividem em classes reunidas pelo grau de semelhança. Além do esquema da BC, Bliss publicou em 1929, a obra “A Organização do conhecimento e o Sistema das Ciências” e em 1933 a obra “A Organização do Conhecimento nas Bibliotecas”, nas quais enfatizava os fundamentos filosóficos da classificação, tentando fazer a combinação entre o conhecimento filosófico e a necessidade pragmática de organizar os livros de uma biblioteca na estante.

Ranganathan buscou inspiração nas ideias de Bliss e criou a *Colon classification* (classificação dos dois pontos ou classificação facetada), cuja principal característica é subdividir os assuntos em facetas. Possui uma notação composta por letras maiúsculas e minúsculas, além de números arábicos, traço de união e dois pontos. Ranganathan trouxe um diferen-

cial para as classificações bibliográficas ao trazer uma terminologia técnica: as facetas.

Os sistemas de classificação bibliográfica já não suportam o volume informacional, visto que o aumento dos acervos e das informações digitais tem criado uma nova realidade, que exige sistemas de classificação mais flexíveis e que correspondam as expectativas no momento da recuperação da informação.

Quadro 2 – Organizações do conhecimento nas CDD e CDU

Classificação Decimal Universal		Classificação Decimal de Dewey	
0	Generalidades	000	Ciências da Computação, Informação e Trabalhos Gerais
1	Filosofia. Psicologia	100	Filosofia e Psicologia
2	Religião	200	Religião
3	Ciências Humanas. Ciências Sociais	300	Ciências Sociais
4		400	Linguística
5	Ciências	500	Ciências
6	Ciências Aplicadas	600	Tecnologia
7	Artes. Desportos	700	Artes e Recreação
8	Línguas. Literatura	800	Literatura
9	Geografia. Biografia. História	900	História e Geografia

Fonte: Própria da pesquisa

A Classificação Decimal de Dewey é baseada no conhecimento e estruturada hierarquicamente em dez classes principais, compostas por disciplinas ou áreas de estudo que são subdivididas. Além da divisão em classes, a CDD conta com tabelas auxiliares de áreas geográficas, períodos históricos, literatura e gêneros literários, grupos raciais e étnicos e uma tabela de línguas. Já a Classificação Decimal Universal, embora seja embasada na CDD, foi mais ambiciosa em seus objetivos, apresentando uma estrutura mais flexível e com uma maior capacidade de síntese e de representação de assuntos mais complexos. Segundo (ANJOS, 2008, p. 179), “[...] permite uma classificação mais detalhada com a incorporação de muitas características de classificação facetada [...]”.

No tocante aos aspectos conceituais da organização do conhecimento, tanto as classificações filosóficas quanto as bibliográficas apresentam seus traços distintivos e suas relações interdisciplinares.

As classificações filosóficas estão relacionadas à hierarquização do conhecimento humano, portanto são mais teóricas, enquanto as classifi-

cações bibliográficas estão ligadas à ordenação dos documentos e referências bibliográficas, sendo práticas.

Enquanto o processo classificatório faz parte da condição humana, a classificação bibliográfica é uma sequência de conceitos planejados para serem aplicados à organização de acervos de bibliotecas para que livros possam ser recuperados de modo eficaz e eficiente, ou seja, uma aplicação pragmática do princípio classificatório no contexto das unidades e sistemas de informação. (ANJOS, 2008, p.132).

Nas classificações filosóficas as particularidades de uma área do conhecimento não são especificadas, já nas classificações bibliográficas as áreas específicas do conhecimento são identificadas e reconhecidas. As classificações bibliográficas fazem um mapeamento da literatura e não da ciência e suas classes são divididas em linhas formais e não filosóficas.

As classificações bibliográficas são adaptações das classificações do conhecimento e frequentemente utilizam termos com origem nas filosóficas, atribuindo novos significados e transformando-os em ferramentas.

As classificações filosóficas e os esquemas de classificação bibliográfica serviram como base para os novos sistemas e estruturas de organização do conhecimento, que tem enfrentado inúmeras dificuldades no acompanhamento da crescente evolução do conhecimento, diante do aumento desenfreado de informações, que não tem sido recuperada com êxito. Surgindo sempre, a necessidade de flexibilidade das classificações a fim de um melhor gerenciamento e organização do conhecimento registrado.

ONTOLOGIAS COMO ESTRUTURAS DE REPRESENTAÇÃO DO CONHECIMENTO NA WEB

A representação do conhecimento é uma área de pesquisa da Inteligência Artificial e segundo Ladeira (p.22, 1997), "Alguns autores definem representação do conhecimento como a formalização e estruturação de conhecimento em um computador, de forma que ele possa ser manipulado pelo sistema gerenciador da base de conhecimento".

Ladeira (1997), afirma que existem três abordagens para se representar o conhecimento: declarativa (ou lógica), procedimental e a estrutural. Na abordagem declarativa, o conhecimento é descrito através de fatos, tendo como exemplo as redes semânticas; na abordagem procedimental, o conhecimento é descrito através de procedimentos para sua utilização, tendo como exemplo regras de produção; e na abordagem estrutural, o conhecimento é descrito com objetos e relações, tendo como exemplo os *frames*.

A representação do conhecimento se dá por meio de linguagens documentais, como os sistemas de classificação, listas de cabeçalho de assunto e tesouros e por meio de produtos documentais, tais como resumos, índices e catálogos.

Na *web* a recuperação da informação é realizada por meio de mecanismos de busca, que se valem da indexação de palavras em linguagem natural, capturando apenas a parte léxica. A quantidade de documentos disponíveis na *web* dificulta a operação dos mecanismos de busca, devido à polissemia inerente à linguagem natural. Os mecanismos de busca também não conseguem recuperar em sua totalidade os documentos disponíveis, visto que, muitos ficam na denominada *web* oculta.

Com o crescimento de usuários e computadores ligados à rede e com a quantidade de informação disponível, torna-se necessário a organização do conteúdo por meio de terminologias padronizadas e controle da linguagem, de forma que a polissemia e a ambiguidade sejam eliminadas. No entanto, uma das principais características do *ciberespaço* é a ausência de uma linguagem padronizada, devido a grande flexibilidade de produção da informação, dificultando a recuperação da informação. A Web Semântica é apresentada como uma solução para a representação dos recursos informacionais na *web*, diminuindo a falta de precisão e os problemas de busca, estruturando a informação e atribuindo sentido.

As estruturas de representação do conhecimento variam de acordo com o ambiente para o qual foram criadas. Os sistemas de classificação, por exemplo, foram criados para as bibliotecas e centros de documentação, já as taxionomias são meios de organizar o conteúdo de *websites*, possibilitando a navegação ilimitada entre os mais variados assuntos por meio de *links*.

As taxionomias podem ser entendidas como um conjunto de termos estruturados tradicionalmente de forma hierárquica, que representam as áreas nas quais são aplicadas. Esse instrumento permite nomear e organizar entidades em grupos que compartilham características similares, tais como correlacionar diferentes linguagens usadas num ambiente, criar mecanismos de acesso, normalizar informações disponibilizadas ou criar palavras-chave e conceitos que categorizam os conteúdos em ramos. (MACULAN et.al. 2009, p.4).

As taxionomias são usadas para pesquisas digitais e se estruturam em categorias e, para serem utilizadas, necessitam de um navegador, seu foco está no usuário, além de ter a vantagem da flexibilidade, diferentemente dos sistemas de classificação. A taxionomia classifica a partir de princípios de semelhanças e diferenças agrupando em classes e utilizando um vocabulário controlado a partir de termos pré-estabelecidos.

Outro instrumento de organização da informação é a *folksonomia*, um conceito novo que surgiu com a propagação da internet e foi criado por Thomas Vander Wal em 2004. Trata-se de uma indexação colaborativa de conteúdos com palavras-chave ou etiquetas de livre escolha, onde se prevalece a linguagem natural. É um instrumento indispensável na indexação, organização e recuperação da informação digital.

A *folksonomia* também é denominada *social tagging*, visto que se trata de uma classificação colaborativa que utiliza palavras-chave, também denominadas *tags*, que são escolhidas pelos usuários de forma livre e utilizadas por eles, conforme seus interesses e necessidades, sem nenhuma hierarquia. Este processo de etiquetagem é realizado de forma espontânea, utilizando uma linguagem natural, onde o usuário escolhe os termos que melhor representem o conteúdo digital, visando facilitar sua posterior recuperação.

A grande vantagem da *folksonomia* é o compartilhamento de informações e a interação entre os usuários, no entanto, devido ao fato de não possuir uma linguagem padronizada e hierárquica, e ter como principal característica a indexação livre, onde os usuários escolhem os termos de acordo com suas preferências, acaba gerando uma grande desorganização devido à ambiguidade e à falta de precisão na recuperação da informação.

A principal vantagem dessa modalidade está no cunho colaborativo social, no qual as pessoas podem livremente participar do processo de indexação, sem que para isso se faça utilizar de regras. Ao contrário, a *folksonomia* não faz distinção dos usuários, quanto ao nível cultural e intelectual de cada um. (ALBUQUERQUE; MESQUITA; COSTA, 2010, p. 5).

As *tags* utilizadas podem apresentar polissemia, sinonímia e erros na grafia e tudo isso impacta de forma negativa na recuperação dos conteúdos. Apesar dessas desvantagens, a *folksonomia* é bastante utilizada como recurso de organização da informação na *web*, e dentre os *sites* que aderiram a esta modalidade estão o *Del.icio.us*, um site que permite que o usuário salve as suas páginas favoritas e o *Flickr*, que é um sistema de publicação de fotos em que os usuários podem indexá-las e estas podem ser visualizadas por outros participantes.

O principal diferencial da *folksonomia* no universo digital é o fato de ser um recurso que permite que os usuários não sejam apenas utilizadores, mas tornem-se construtores da informação.

Outro recurso bastante utilizado são as redes semânticas, uma forma de representação do conhecimento que utilizam nós, ligações e rótulos. Os nós representam categorias, propriedades ou objetos, as ligações de-

finem as relações e os rótulos identificam as ligações. As redes semânticas são uma forma mais flexível de representação do conhecimento que possibilita relações entre classes e instâncias.

Uma evolução das redes semânticas são os sistemas de *frames*, estruturas de dados para representação do conhecimento, organizado de forma hierárquica e que ao invés de nós, utiliza *frames* para representar as categorias, possuindo pelo menos dois atributos e dois tipos de relações estruturais entre classe e subclasse e entre classe e instância.

Assim como nas redes semânticas, uma das características dos frames é a possibilidade de se definir novos subtipos de objetos que herdam todas as propriedades da classe original. A diferença básica das redes semânticas em relação aos frames é que, enquanto o primeiro organiza o conhecimento na forma de redes, o segundo organiza de forma hierárquica [...]. (SERGIO FURGERI, 2006, p. 30)

Tanto as redes quanto os *frames* semânticos, apesar de apresentarem uma estrutura capaz de possibilitar a descrição de objetos de domínio e suas relações, apresentam algumas limitações e aspectos negativos como a baixa precisão podendo gerar ambiguidade, inferências incorretas e, no caso dos *frames*, apresenta suporte apenas para a estrutura hierárquica, o que pode gerar restrições de representação.

Outra estrutura de representação do conhecimento são as Ontologias, não no seu sentido filosófico de estudo do ser e da existência, mas como um instrumento capaz de atribuir semântica ao conteúdo dos documentos.

Uma Ontologia permite o compartilhamento de um domínio tornando-se, portanto, indispensável no desenvolvimento da Web Semântica (Figura 1), permitindo o acesso e a interoperabilidade de conteúdo. Além de apresentar uma estrutura flexível que pode ser redefinida ou adaptada dependendo do contexto no qual está inserida, permitindo a extensão de suas relações.

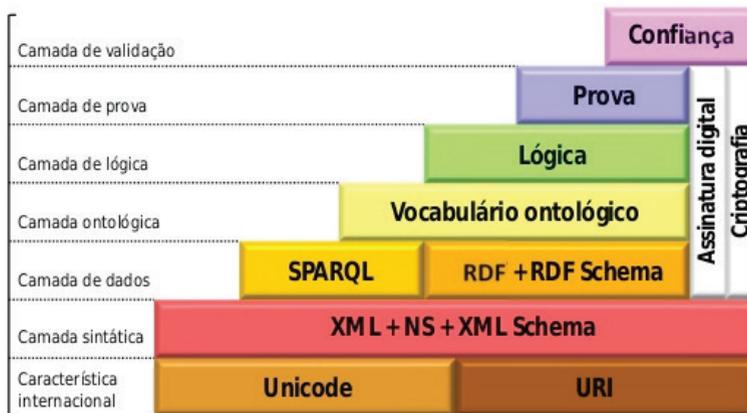
Uma das mais fortes razões para o desenvolvimento de ontologias é a possibilidade de compartilhamento e reutilização de conhecimento formalmente representado para uso em sistemas de Inteligência Artificial e de Arquitetura da Informação, o que exige a definição de um vocabulário comum para a Representação do Conhecimento a ser aplicado [...]. (LIMA-MARQUES, p. 52, 2006).

Normalmente uma Ontologia é especificada por uma taxonomia para definir as classes e subclasses e suas relações; e por um conjunto de regras de inferência capazes de realizar deduções e manipular termos por meio de um programa de computador. As Ontologias surgem como uma

resposta na melhoria da precisão das buscas de dados na *web*, de forma que problemas tais como, ambiguidade e polissemia sejam minimizados.

As Ontologias representam algumas definições básicas dentro de um domínio específico, onde conceitos e suas relações são legíveis por um computador, de forma que esse conhecimento possa ser reutilizado e compartilhado por outros sistemas e pessoas.

Figura 1 – Estrutura da Web semântica



Fonte: Adaptado do livro *A Web Semantic Primer*

A Ontologia é uma das tecnologias utilizadas na Web Semântica, e segundo Vieira e Freitas (2008), é estruturada em camadas (figura 1). A camada da base *Unicode* e *URI* é a responsável pelo fornecimento da universalidade de acesso a informações relacionadas à codificação de caracteres, endereçamento e nomeação de recursos. A camada *XML + Namespace + xmlschema* fornece a interoperabilidade em relação à sintaxe. Já a camada *RDF + rdfshema* fornece a interoperabilidade em relação à semântica. A camada Ontologia fornece suporte para a evolução e especificação de conceitos e vocabulários, enquanto a camada Lógica permite o relacionamento lógico entre os conceitos definidos na camada Ontologia. As camadas de Prova e Confiança dão o suporte para a execução das regras, à medida que avaliam e corrigem todo esse processo.

Técnicas e tecnologias de desenvolvimento e implementação de Ontologia

A Ontologia busca representar o conhecimento humano de forma estruturada, no sentido filosófico, representada a partir de uma visão particular de mundo.

No âmbito da Ciência da Informação, o termo ontologia identifica-se com a definição dada a ele na Ciência da Computação, ou seja, uma ontologia denota as especificações dos conceitos relevantes para um assunto no domínio, incluindo suas propriedades e relações entre eles, e uma classificação hierárquica destes conceitos em categorias e subcategorias. (LIMA, p.11, 2013).

Além da função de uma linguagem documentária, as Ontologias também são consideradas um instrumento de organização do conhecimento que visa diminuir as ambiguidades da linguagem natural, utilizando para tal um padrão terminológico por meio de um vocabulário específico capaz de representar uma realidade.

Segundo Rautenberg et. al. (2008), a terminologia de Engenharia de Ontologias, visto que é baseada na engenharia de *software*, deve considerar algumas atividades durante seu processo de criação: a **especificação**, que tem a tarefa de identificar o propósito e o escopo da Ontologia; a **conceitualização**, cuja tarefa é construir o modelo conceitual da Ontologia; a **formalização**, que transforma a descrição conceitual em um modelo formal; a **implementação**, que transforma a Ontologia formalizada em uma linguagem de representação do conhecimento; e a **manutenção**, responsável por fazer as atualizações e correções na Ontologia desenvolvida.

Já as representações nas Ontologias são de duas formas: a **terminológica**, onde se usa as relações de categoria para definir os termos da linguagem natural; e a **formal**, em que se usa uma linguagem própria que permite o processamento por máquina. O principal objetivo ao se construir uma Ontologia é a necessidade de um vocabulário compartilhado. Na construção de Ontologias devem-se observar alguns princípios-chaves como: clareza, legibilidade, coerência e extensibilidade.

As Ontologias podem ser classificadas quanto ao grau de formalismo em: altamente informal, com uma linguagem natural desestruturada; semi-informal, com uma linguagem natural estruturada e restrita; semi-formal, com uma linguagem artificial definida; e rigorosamente formal, com termos bem definidos através de semântica formal.

Quanto à natureza podem ser classificadas em: Ontologias de Tarefa e Ontologias de Domínio.

As Ontologias de Tarefa são aquelas que fornecem um vocabulário sistematizado de termos, especificando tarefas genéricas que podem ou não estar no mesmo domínio [...] descrevem atividades através de um vocabulário formado por ações com a finalidade de possibilitar a execução de tarefas pela máquina. (SALES, 2006, p.32)

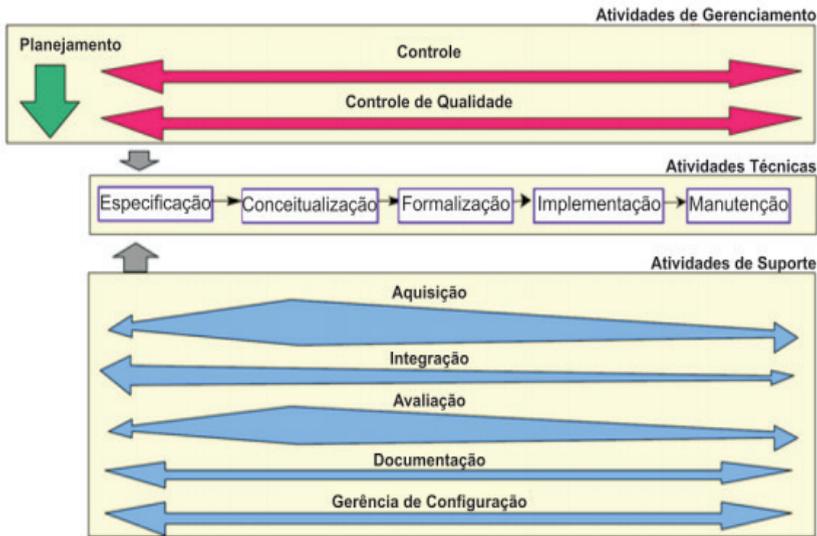
Já as Ontologias de Domínio são reutilizáveis em um domínio genérico, fornecendo vocabulários sobre conceitos, seus relacionamento, atividades e regras que os governam. “[...] é próprio para representação e recuperação das informações de um dado domínio de conhecimento.” (SALES, p.32).

No processo de construção de Ontologias de Domínio pressupõe-se a definição de seu domínio e escopo, antes de se escolher a metodologia, ferramenta e linguagem para sua especificação. Embora nem todas as Ontologias possuam a mesma estrutura Sales (2006 apud TELLO, 2004) afirma que, uma Ontologia apresenta os seguintes elementos em sua estrutura: **conceitos**, que são ideias básicas sobre o que se pretende formalizar; **classes e subclasses**, que podem formar uma taxonomia; **relações**, que representam os tipos das relações entre as classes de domínio; **funções**, que são relações especiais que não permitem a ordem inversa; **axiomas**, que são os teoremas que se declaram sobre as relações; e **instâncias**, que são utilizadas na representação de objetos de um conceito.

Morais e Ambrósio (2007), afirmam que apesar de não existir metodologias suficientemente maduras para a construção de Ontologias, podem-se citar as seguintes:

- *Enterprise* – construção baseada em quatro fases: identificação do propósito, identificação do escopo, formalização e documentação formal;
- *Methontology* – construção baseada a partir de um conhecimento de um domínio;
- *On-To-Knowledge* – construção baseada em quatro fases: *kick-off*, refinamento, avaliação e manutenção.

No processo de construção de Ontologias, a metodologia mais utilizada é a *Methontology* (figura 2), idealizada por um grupo de pesquisa da Universidade Politécnica de Madri e segundo Rautenberg et.al. (2008), a característica principal dessa metodologia é a definição de um rico conjunto de artefatos de documentação, cujo ciclo de vida se baseia na prototipagem de Ontologias. Seu processo de criação de Ontologias apresenta semelhanças com a criação de um *software*, a construção da Ontologia é baseada a partir do conhecimento de um domínio. “Suas atividades principais são: especificação de requisitos, conceitualização do domínio do conhecimento, formalização do modelo conceitual em uma linguagem formal, implementação de um modelo formal e manutenção de ontologias implementadas.” (PAULA, 2014, p.60).

Figura 2: Processo de construção de Ontologias com a *methontology*

Fonte: Paula (2014)

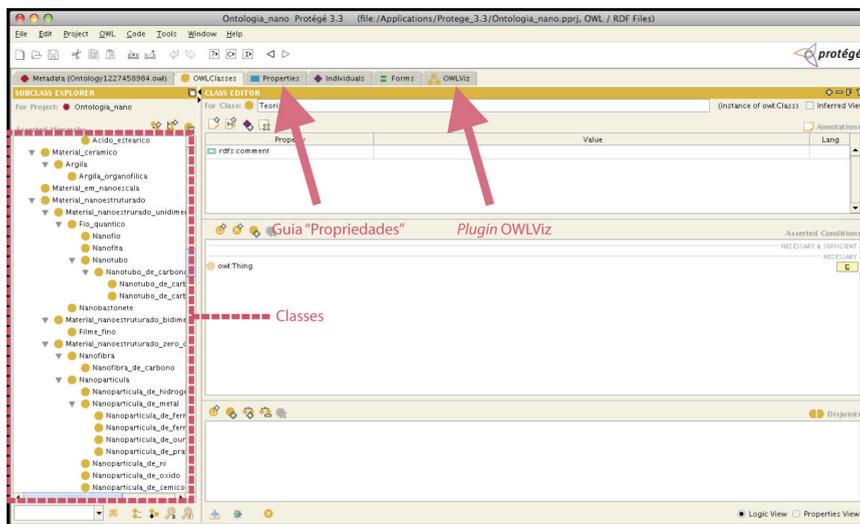
As ferramentas utilizadas para construção de Ontologias são inúmeras, como por exemplo, o **WebODE**, uma aplicação *web*, que armazena Ontologias em bases de dados relacionais, desenvolvida no laboratório de inteligência artificial da Universidade Politécnica de Madri e com suporte para *Methontology*; o **OntoEdit** “[...] um ambiente gráfico para edição de ontologias que permite inspeção, navegação, codificação e alteração de ontologias [...]” e possui suporte para *On-To-Knowledge*. (MORAIS; AMBRÓSIO, 2007, p.15).

Outra ferramenta desenvolvida e coordenada por membros da Universidade Federal de Santa Catarina (UFSC) é a **OntoKEM**, que além de especificar, conceituar, formalizar e documentar Ontologias, gera automaticamente artefatos customizados para documentar projetos de Ontologias. Porém, Rautenberg; Todesco; Gauthier (2009), afirmam que, a ferramenta tem a limitação de não poder participar de todo o processo de desenvolvimento de Ontologias, como das atividades de implementação e validação, sendo complementado nessa tarefa pelo **Protégé**.

O **Protégé** (figura 3) é de uso gratuito “[...] é fundamentado em frames e pode ser carregado em qualquer computador [...]” (LIMA-MARQUES, p.62). É um editor de código aberto desenvolvido em Java na Universidade de Stanford, utilizado nas mais diversas áreas para recuperação

da informação e modelagem corporativa, além de suportar dois tipos de modelagem de Ontologias *Frame* e OWL¹.

Figura 3: Editor de Ontologia Protégé



Fonte: Kasama (2009)

O **Protégé** permite construir Ontologias, personalizar fórmulas de entrada de dados e inserir e editar dados. Além de refinar o projeto de Ontologias na fase de implementação, propiciando o ambiente de teste para a Ontologia construída. Sua interface gráfica permite o acesso à barra de menus e ferramentas, além de apresentar áreas de visualização (*views*) que permitem a navegação e a edição de classes, atributos, formulários e pesquisas dentro de toda a base de conhecimento.

Seu projeto original foi desenvolvido para um sistema especialista em oncologia e, ao longo do tempo foi modernizado, a fim de permitir os mais diversos formalismos e estratégias de inferência, integrando tarefas em um mesmo ambiente, além de criar automaticamente formulários para entrada de conhecimento, bem como o acesso e combinação de Ontologias.

As linguagens para representação de Ontologias segundo Almeida e Bax (2003), podem ser divididas em três tipos: linguagens de Ontologias tradicionais, linguagens padrão *web* e linguagens de Ontologias *Web-based*. No entanto, o critério para escolha varia de acordo com o tipo

¹ OWL (*Ontology Web Language*) - linguagem padrão para a construção de Ontologias desenvolvida para atender as necessidades das aplicações da Web Semântica, tendo com elementos básicos *namespaces*, cabeçalhos, classes, indivíduos, propriedades e restrições. (VIEIRA E FREITAS, 2008).

de Ontologia a ser especificado, as linguagens recomendadas pelo W3C (*World Wide Web Consortium*) são a Ontolândia, RDF e OWL. A Ontolândia combina paradigmas de *frames* e predicados de primeira ordem, além de permitir a representação de conceitos, taxonomia de conceitos, axiomas, instâncias e procedimentos. O RDF é uma linguagem baseada em rede semântica que permite a representação de conceitos, taxonomias de conceitos e relações binárias. E a OWL é utilizada para representar o significado de termos em vocabulários e os relacionamentos entre estes.

Atualmente a linguagem mais utilizada na representação de Ontologias é a XML (*Extensible Markup Language*), devido a sua popularidade no ambiente da *web*, tanto na construção de arquivos quanto no intercâmbio de dados. No entanto, a linguagem recomendada pela W3C para a construção de Ontologias é a OWL, devido a ser compatível com a arquitetura da *World Wide Web* e da Web Semântica, e permitir a distribuição de uma Ontologia através de vários sistemas.

Segundo Carvalheira (2007), a OWL apresenta três sub-linguagens:

- OWL *Lite*: com suporte para representação de Ontologias mais simples, com hierarquia de classificação e restrições menos complexas;
- OWL DL: com maior poder de expressividade, porém com algumas restrições como a impossibilidade de uma classe não poder ser instância de outra classe;
- OWL *Full*: com alto poder de expressividade, no entanto, tem seu limitado por mecanismos de raciocínio automáticos.

A linguagem OWL permite a organização hierárquica em classes e subclasses, declarando propriedades e sub-propriedades, a expressão dos objetos e as classes as quais pertencem, além de permitir que as declarações de equivalência possam ser feitas para classes ou propriedades. Essa possibilidade de inclusão de um maior número de vocabulários na descrição das classes e propriedades gera um domínio definido com maior expressividade.

Além de construídas, as Ontologias podem ser alinhadas com outras já existentes, desde que sejam de domínios complementares, que apesar de serem de assuntos diferentes apresentem partes em comum. Neste processo, duas Ontologias originais separadas, por apresentarem termos semelhantes podem ter seus termos ligados, permitindo assim que as Ontologias alinhadas reusem as informações umas das outras.

Há também outros processos segundo Vieira e Freitas (2008), como o mapeamento de Ontologias, no qual se faz apenas uma análise de se-

melhanças, sem nenhum tipo de modificação das Ontologias envolvidas. A combinação que gera uma Ontologia única, com os termos juntos e o processo de integração, no qual a Ontologia produzida é caracterizada pela extensão, especialização ou adaptação de outras Ontologias que re-metem a assuntos diferentes.

Domínios de Ontologias representados

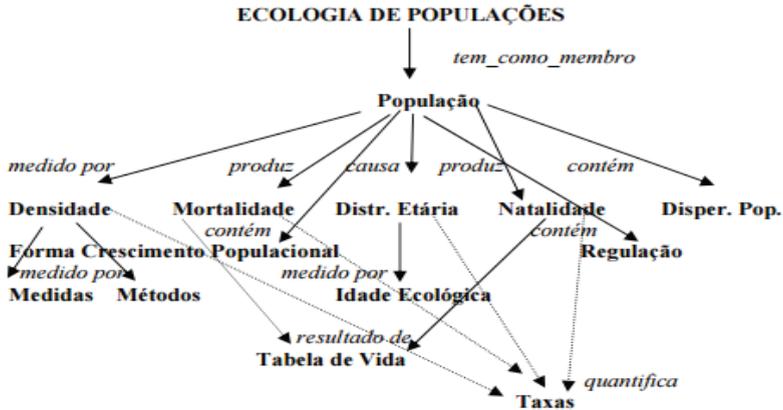
Para Sales e Café (2008), “As ontologias, que são aparatos desenvolvidos pela engenharia computacional, consistem em especificações formais que visam descrever estruturas conceituais de domínios específicos”. Por meio de uma Ontologia pode-se fazer a representação de um conjunto de conceitos e seus relacionamentos dentro de um domínio específico. Trazendo como proposta a captura de domínios do conhecimento, fornecendo um entendimento semântico e permitindo o compartilhamento de tais informações.

A base gerada por meio de seus conceitos chaves permite, além da análise do conhecimento do domínio, a reutilização do conhecimento de domínios específicos e a construção de outros domínios. Neste estudo trago alguns domínios de Ontologias em algumas áreas como Ciências Biológicas, Nanociência e Nanotecnologia, Direito, Esportes e Cultura, a fim de mostrar a diversidade de possibilidades de criação de Ontologias nos mais variados domínios.

- **OntoEco – Domínio das Ciências Biológicas – Subdomínio da Ecologia**

A OntoEco foi desenvolvida utilizando o Protegé 2000 e prevê três subdomínios da Ecologia: Ecologia de Ecossistemas; Ecologia de Populações (figura 4); e Ecologia de Comunidades, que durante os processos de categorização e nomeação de termos se mostraram bastante produtivos. Segundo Claudia Zavaglia (2002), a OntoEco encontra-se dividida em duas classes: CLASSES E LEXICAL_UNIT. Estas se subdividem em subclasses, a primeira em classes de semântica-base, que define o padrão de configuração das classes e subclasses vinculadas a elas; e a última tem o padrão definido por meio da classe de unidade lexical-base. Os seus termos possuem a relação *é_um* e *é_um_tipo_de*, a relação *é_um* determina todas as subclasses das classes principais CLASSES E LEXICAL_UNIT.

Figura 4: Ecologia de Populações - relacionamentos



Fonte: Zavaglia (2003)

• Domínio da Nanociência e Nanotecnologia

O portal OntoLP (<http://www.inf.pucrs.br/ontolp/downloads.php>) disponibiliza uma base de dados de algumas Ontologias, dentre elas está o projeto desenvolvido pela coordenadora Sandra Maria Aluísio, no qual foi criado um *corpus* sobre nanotecnologia. O projeto foi baseado totalmente na língua inglesa e desenvolvido visando à organização do Portal da Rede de Nanotecnologia da USP (Universidade de São Paulo), foi elaborado um *corpus* com a extensão de 2.570.792 palavras e uma Ontologia contendo cerca de 1.900 termos.

Esta pesquisa estimulou outra, coordenada pela Profa. Dra. Gladis Maria de Barcellos da Universidade Federal de São Carlos que tinha como objetivo a sistematização para língua portuguesa. Conforme Kasama (2009), um *corpus* constituído por 2.565.790 palavras foi gerado, divididas entre os gêneros Científico, Informativo e Científico de Divulgação, Técnico-Administrativo e outros. E como resultado gerou um domínio com 6 subdomínios : **Aplicações** (figura 5) , **Equipamentos**, **Materiais**, **Métodos e Técnicas**, **Propriedades e Teorias**, totalizando 361 classes e subclasses.

Figura 5: Subdomínio Aplicações



Fonte: Kasama (2009)

O subdomínio Aplicações está dividido em 6 classes (Biologia, Computação, Eletrônica, Física, Medicina e Química) e possui um total de 17 classes e subclasses.

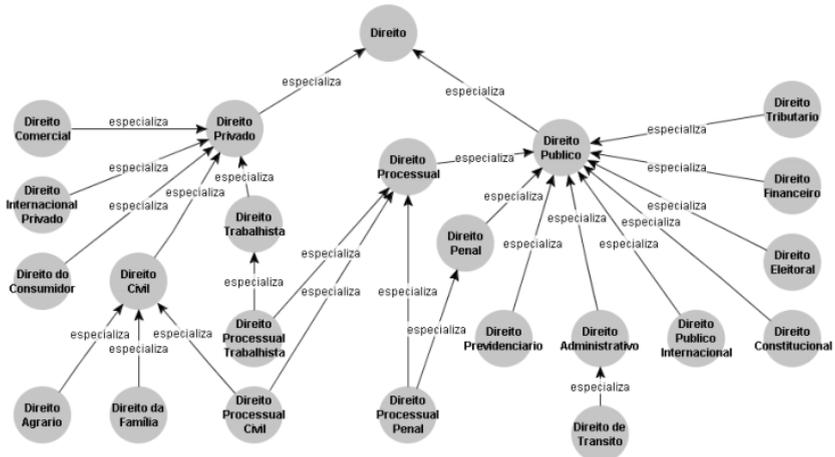
• Ontojuris – Domínio Jurídico – Direito Eletrônico

O projeto Ontojuris (<http://www.i3g.org.br/ontojuris>) é realizado pela Universidade Federal de Santa Catarina sob a coordenação do professor Aires José Rover Júnior e envolve pesquisadores do Brasil, Chile, Espanha e Argentina, visa facilitar o acesso a informações sobre legislação na área de Propriedade Intelectual, Direito do Consumidor e Direito Eletrônico. Segundo Júnior Rover (2009), a Ontologia “[...] se propôs a ajudar a esclarecer conceitos para este domínio [...] auxiliando no esclarecimento da tipicidade de condutas criminosas e suas características e na recuperação de obras científicas e jurisprudências sobre estes delitos”.

A Ontojuris (figura 6) foi desenvolvida com a finalidade de modelar o conhecimento jurídico brasileiro e para tal, dividiu o Direito em ramos

específicos permitindo a relação entre cada instrumento normativo e sua respectiva área.

Figura 6: Hierarquia dos ramos jurídicos da Ontojuris



Fonte: Moron (2010)

A Ontojuris facilita o acesso a informações sobre legislação e para tal conta com uma estrutura computacional representada pela tecnologia de buscas inteligentes do instituto i3G, cuja busca é baseada em Ontologias, que permite o mapeamento do assunto e suas relações de forma contextualizada. Cabe destacar que a Ontojuris possui uma extensão denominada Ontotrib que é uma Ontologia especializada em direito tributário.

• RANWEZ – Domínio do Esporte – Subdomínio do Futebol

A RANWEZ é uma Ontologia do futebol proposta por Ranwez em 2002, que descreve conceitos específicos do futebol, tais como: jogadores, regras, campo, patrocinadores e etc. Segundo Faria (2009), “O objetivo principal desta ontologia era para ser usada para anotação de vídeos com o propósito de serem produzidos sumários personalizados dos jogos de futebol.” No entanto, por sua estrutura não possuir as classes organizadas hierarquicamente (figura 7) e nem ter relação entre os conceitos, pode ser classificada como um vocabulário controlado.

Figura 7: Hierarquia de classes estruturais



Fonte: Faria (2009)

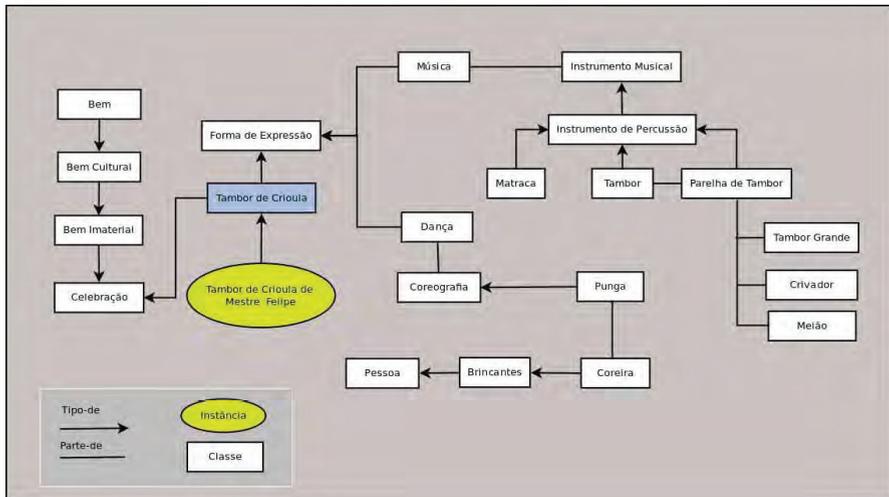
Por meio desta Ontologia foi desenvolvida a *SWAN Soccer Ontology*, também considerada um vocabulário controlado, porém, apresenta suas classes já organizadas com uma hierarquia simples, com os termos organizados por tipo.

- **DCPCSLs – Domínio Conceitual “Patrimônio Cultural de São Luís do Maranhão”**

O DCPCSLs (figura 8) foi desenvolvido pela Dr. Valdirene Pereira da Conceição e tem como suporte referencial o patrimônio cultural da cidade de São Luís, com uma modelagem que apresenta uma versão exploratória do domínio.

O *corpus* da Ontologia foi extraído de textos da literatura especializada sobre patrimônio cultural e foi formalizada e implementada no editor Protegé. O DCPCSLC é constituído por seis classes: **Bem Cultural; Valor; Formas de Proteção; Uso; Órgãos de Preservação; e Classificação**. Gerou uma Ontologia composta de 256 unidades ontológicas.

Figura 8: Fragmento da Ontologia do DCPCSLs



Fonte: Conceição (2011)

Entre as grandes categorias conceituais do DCPC estão: o tipo de bem cultural apresentado; a natureza desse bem cultural; o valor que ele possui para a comunidade; a legislação/política de preservação/conservação; as medidas de proteção; e as referências feitas à identidade, à história e à memória.

A produção científica sobre Ontologias e a representação nas bases de dados

As bases de dados são repositórios de publicações armazenados em plataformas ou portais, sustentados por tecnologias da informação. A produção científica de Ontologias tem crescido bastante nos últimos dias e é grande a disponibilidade de textos nas bases de dados, que reúnem informações relevantes nas mais variadas áreas do conhecimento. Nesta pesquisa analisaram-se as seguintes bases de dados: CAPES - LILACS - SCIELO - BDTD. Em cada uma delas, foi feito o mapeamento da quantidade de ocorrências do termo Ontologia, bem como o gênero documental e a característica da temática do termo, se tratava de simples conceito ou metodologia para desenvolvimento.

Quadro 3 - Produção Científica de Ontologias nas Bases de Dados

BASE DE DADOS	OCORRÊNCIAS	ENDEREÇO
CAPEB	1.399	http://www.periodicos.capes.gov.br/
LILACS	138	http://pesquisa.bvsalud.org/porta//
SCIELO	345	http://www.scielo.org/php/index.php
BDTD	1.220	http://bdtb.ibtct.br/vufind/

Fonte: Própria da pesquisa

O portal de periódicos CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) é responsável pela centralização e otimização da aquisição do conteúdo das bibliotecas virtuais, por meio da negociação direta com os editores internacionais. Na pesquisa foram encontrados 1.399 ocorrências em 15 idiomas: espanhol, português, inglês, francês, italiano, alemão, catalão, polonês, romeno, húngaro, russo, chinês, galego, italian e basco.

Quadro 4 - Análise do portal de periódicos CAPES

TÍTULO	CATEGORIA	PERIÓDICO
A Guide to Understanding Social Science Research for Natural Scientists	Sentido Filosófico	Conservation Biology, 2014, Vol.28(5), pp.1167-1177
Ontología de evaluación de impacto ambiental para proyectos mineros	Ontologia/Domínio da Mineração	Minería y Geología, 01 March 2014, Vol.30(1), pp.104-116
Ontología de parlamentarios chilenos autores: Modelamiento y aplicación	Ontologia/Domínio de Autores Parlamentares	Serie Bibliotecología y Gestión de Información, 2014, Issue 91, p.
La ingeniería del conocimiento en el dominio legal: La construcción de una Ontología Satélite en FunGramKB	Ontologia/Domínio Jurídico	Revista signos - Estudios de lingüística, 2014, Vol.47(84), pp.113-139
El método de Freud y la tradición hermenéutica en psicoanálisis	Sentido Filosófico	Revista Latinoamericana de Metodología de las Ciencias Sociales, 2014, Vol.4(2), pp.34-61
"Study about xprw and jrw and dynamics in the funerary literature of the royals Theban tombs. An approach to the ontological mobility in the Duat"	Sentido Filosófico	Sociedades Precapitalistas, 01 June 2015, Vol.4(2)
El papel de los afectos en el pensamiento político de Spinoza	Sentido Filosófico	Ideas y valores: Revista Colombiana de Filosofía, 2014, Issue 154, pp.31-57
Hacia una filosofía materialista: : La idea de Naturgeschichte en la obra de Theodor Adorno	Sentido Filosófico	Revista de filosofía, 2014, Issue 70, pp.75-87
Marcelo Cohen. Las fundaciones de la ciencia ficción (parte II: La comunidad)	Sentido Filosófico	Orbis Tertius, 01 June 2015, Vol.20(21), pp.30-38
Marcelo Cohen. Las fundaciones de la ciencia ficción	Sentido Filosófico	Orbis Tertius, 01 February 2014, Vol.18(19), pp.79-89

Hacia una filosofía materialista: la idea de Naturgeschichte en la obra de Theodor W. Adorno Towards a materialistic philosophy: the idea of naturphilosophie in the work of Theodor W. Adorno	Sentido Filosófico	Revista de filosofía, 2014, Vol.70(), pp.75-87
Estados de cosas negativos: La perspectiva fenomenológica de Adolf Reinach On Negative States of Affairs: The Phenomenological Perspective of Adolf Reinach	Sentido Filosófico	Revista de filosofía y teología, 2014, Issue (30), pp.65-81
La noción de ousia en el didaskalikós de alcinoo	Sentido Filosófico	Revista de Filosofía y Teoría Política, 2014, Issue 45
Linguistic Anthropology in 2013: Super!New!Big	Sentido Filosófico	American Anthropologist, 2014, Vol.116(2), pp.366-378
Pensar topográficamente: Lugar, espacio y geografía	Sentido Filosófico	Documents d'anàlisi geogràfica, 2015, Vol.61(2), pp.199-229
Ontology, Epistemology, Strategy and Method in Educational Research: A Critical Realist Approach	Sentido Filosófico	Revista Internacional de Investigación en Educación, 2014, Vol.7(14), pp.29-38
Los supuestos que subyacen a las principales teorías de Ann Wilcock y la necesidad de la pregunta ontológica por la ocupación humana.	Sentido Filosófico	Revista electrónica de terapia ocupacional Galicia, TOG, 2015, Issue 21
Desarrollo de un Sistema de Recuperación de Información para Publicaciones Científicas del Área de Ciencias de la Computación	Ontologias/Domínio da Ciência da Computação	Revista Latinoamericana de Ingeniería de Software, 01 April 2014, Vol.2(2), pp.107-114

Fonte: Própria da pesquisa

Ao se fazer a busca de publicações do período de 2014 a 2015, refinada com o tópico Ontologia obteve-se 18 ocorrências, conforme o quadro 4, dos 18 artigos disponíveis, 4 são sobre Ontologias nos seguintes domínios: Mineração, Autores Parlamentares, Jurídico e Ciência da Computação. Os outros 14 artigos encontrados apresentam o termo Ontologia no seu sentido filosófico.

A base de dados LILACS (Literatura Latino-Americana e do Caribe em Ciências da Saúde) é uma base cooperativa do sistema BIREME², que indexa os mais variados tipos de documentos como artigos de revistas, teses, livros, anais de congressos, relatórios técnico-científicos e publicações governamentais relativos à área da saúde e produzidos por autores latino-americanos. A pesquisa apresentou 138 ocorrências do termo Ontologia, sendo 121 artigos, 13 teses e 4 monografias, sendo encontradas em três idiomas: português, espanhol e inglês.

2 BIREME – Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde, também conhecido pelo seu nome original Biblioteca Regional de Medicina.

Quadro 5 - Análise da base de dados LILACS

TÍTULO	CATEGORIA	PERIÓDICO
Ontologias para a Modelagem Multiagente de Sistemas Complexos em Ciências Cognitivas	Sentido Filosófico	Ciênc. cogn; 19(1): 58-75, mar. 2014. ilus.
Transduction motif analysis of gastris câncer based on a human signaling network	Ontologia/Domínio da Genética	Braz J Med Biol Res; 47(5): 369-375, 02/05/2014. tab, graf.
Alignment of ICNP Ontology and a proposed INCP Brazilian Ontology	Ontologia/Domínio da Enfermagem	Rev Lat Am Enfermagem; 22(3): 499-503, May-Jun/2014
Expression analysis of miRNA and target mRNAs in esophageal cancer	Ontologia/Domínio da Genética	Braz J Med Biol Res; 47(9): 811-817, 09/2014. tab, graf.
Bioinformatics analysis of biomarkers and transcriptional factor motifs in Down syndrome	Ontologia/Domínio da Genética	Braz J Med Biol Res; 47(10): 834-841, 10/2014. tab, graf.
Sedea: Prototipo de Sistema expert para el diagnóstico de enfermedades autoinmunes de órgano basado em internist	Ontologia/Domínio da Imunologia	Univ. salud; 16(2): 207-218, jul.-dic. 2014. ilus, tab.
Rutas ontológicas de la nosología psiquiátrica: ¿Cómo se llegó al DSM-5? /	Sentido Filosófico	Rev Neuropsiquiatr; 78(1): 35-45, ene. 2015.
A validation of the constructo and reliability of na emotional intelligence scale applied to nursing students	Ontologia/Domínio da Genética	Rev Lat Am Enfermagem; 23(1): 139-147, Jan-Feb/2015. tab
Lacaziosis – unusual clinical presentation	Ontologia/Domínio da Genética	An Bras Dermatol; 90(2): 268-269, Mar-Apr/2015. graf.
Terminologies and classification systems in biomedical sciences	Ontologia/Domínio da Biomedicina	Rev. cuba. inform. méd; 7(1): 89-104, ene.-jun. 2015.
Articulação entre hermenêutica e fenomenologia para uma clínica psicológica com bases heideggerianas	Sentido Filosófico	Rev. abordagem gestál. (Impr.); 21(1): 74-82, jun. 2015.
Ética, saúde coletiva, pesquisa qualitativa em saúde e justiça social	Sentido Filosófico	Ciênc. saúde coletiva; 20(9): 2631-2640, Sep. 2015.

Fonte: Própria da pesquisa

Ao se fazer a busca de publicações do período de 2014 a 2015, refinada com o tópico Ontologia obteve-se 12 ocorrências, conforme o quadro 5, dos 12 artigos disponíveis, 8 são sobre Ontologias nos seguintes domínios: Genética, Enfermagem, Imunologia e Biomedicina. Os outros 4 artigos encontrados apresentam o termo Ontologia no seu sentido filosófico.

A SCIELO (Scientific Eletronic Library Online) é uma biblioteca eletrônica de periódicos científicos brasileiros, que visa aumentar a visibilidade, acessibilidade e qualidade dos mesmos. O programa é mantido pela FAPESP (Fundação de Amparo à Pesquisa do estado de São Paulo) e conta com o apoio do CNPq. A pesquisa apresentou 345 artigos em cinco idiomas: espanhol, português, inglês, italiano e francês.

Quadro 6 - Análise da base de dados SCIELO

TÍTULO	CATEGORIA	PERIÓDICO
Alignment of ICNP Ontology and a proposed INCP Brazilian Ontology	Ontologia/Domínio de Enfermagem	Rev. Latino-Am. Enfermagem; 22(3); 499-503; 2014-06
Por uma ontologia plural de vozes singulares: o embate de Adrian Cavarero com a metafísica	Sentido Filosófico	Rev. Estud. Fem.; 23(1); 282-284; 2015-04
Los filósofos estoicos: Ontología, Lógica, Física y Ética	Sentido Filosófico	Tópicos (México); (47); 235-240; 2014-12
Ontología y Lenguaje de la Realidad Social	Sentido Filosófico	Cinta moebio; (50); 70-79; 2014-09
Aprendizaje ontológico para el marco ontológico dinámico semântico	Sentido Filosófico	Dyna rev.fac.nac.minas; 81(187); 56-63; 2014-10
El principio de inmanencia y la diferencia ontológica	Sentido Filosófico	Tóp. Sem; (31); 175-193; 2014-06
A ruptura articulatória dos seres: a propósito da exposição da vida à dispersão da sua ontologia	Sentido Filosófico	Sci. stud.; 12(2); 359-377; 2014-06
O cuidado de si e os perigos de uma ontologia ainda sem cabimento: o legado ético-espiritual de Foucault	Sentido Filosófico	Pro-Posições; 25(2); 121-138; 2014-08
La ingeniería del conocimiento en el dominio legal: La construcción de una Ontología Satélite en FunGramKB	Ontologia/Domínio Jurídico	Rev. signos; 47(84); 113-139; 2014-03
A reference ontology for harmonizing process-reference models	Ontologia/Domínio de Referência de Modelos de Processos	Rev.fac.ing.univ. Antioquia; (73); 29-42; 2014-12
Dos alternativas de ontología angélica	Ontologia/Domínio de Anjos	Cuest. teol.; 41(95); 75-96; 2014-01
Propuesta para extender semanticamente el proceso de recuperación de información	Desenvolvimento e implementação de Ontologias	Rev.EIA.Esc.Ing.Antioq; (22); 51-65; 2014-12
La semilla del humanism Moralidad y ontología fundamental em Heidegger	Sentido Filosófico	Co-herencia; 11(20); 121-139; 2014-06
Identifying dead features and their causes in product line models: na ontological approach	Ontologia/Domínio das causas de morte	Dyna rev.fac.nac.minas; 81(183); 68-77; 2014-02
La especularidad de la carne: Sobre el sentido del "giro ontológico" em Le Visible el l'invisible de Merleau-Ponty	Sentido Filosófico	Diánoia; 59(72); 85-111; 2014
Capitalismo e ontologia da bioética: o processo de trabalho de médicos de família na Itália	Sentido Filosófico	Trab. educ. saúde; (ahead); 00-00; 2015
Discurso militar e identidade nacional chilena	Sentido Filosófico	Polis; 13(38); 467-498; 2014-08

Uma abordagem integrada sobre ontologias: Ciência da informação, Ciência da Computação e Filosofia	Perspectivas sobre o termo Ontologia	Perspect. ciênc. inf.; 19(3); 242-258; 2014-09
Creación y uso de una ontología relacionada con genes, síndromes, síntomas y enfermedades para la clasificación de textos biomédicos	Ontologia/ Domínio da Genética	Rev. signos; 47(84); 91-112; 2014-03
Capitalismo e ontologia da bioética: o processo de trabalho de médicos de família na Itália	Sentido Filosófico	Trab. educ. saúde; 13(2); 323-341; 2015-08
Atención, referencia e inescrutabilidad	Sentido Filosófico	Estud.filos; (50); 31-51; 2014-12
Entre a epistemologia e a ontologia: a teoria da estruturação de Anthony Giggens	Sentido Filosófico	Tempo soc.; 26(2); 123-136; 2014-12
Hacia una filosofía materialista: la idea de Naturgeschichte em la obra de Theodor W. Adorno	Sentido Filosófico	Rev. filos.; 70(); 75-87; 2014-11-01
Pensar desde el cuerpo: de Merleau-Pünty a Jean-Luc Nancy y el nuevo realismo	Sentido Filosófico	Eidos; (21); 221-236; 2014-07-15
Poiesis del tempo y del movimiento: una nueva mirada a la ontología aristotélica	Sentido Filosófico	Univ. philos.; 31(63); 267-286; 2014-12
(Trans)formación sociotécnica, subjetividad y política	Sentido Filosófico	Rev. de Inv. Educ.; 7(3); 34-51; 2014-12
A imagem e as ciências humanas: a poética visual como possibilidade de construção do saber	Sentido Filosófico	Educ. Real.; 39(4); 1153-1168; 2014-12
Representación del Conocimiento de la Información Geográfica siguiendo um Enfoque basado em Ontologías	Ontologia de representação/ Domínio Geoespacial	RISTI; (14); 101-116; 2014-12
La philia como investigación fenomenológica particular del saber práctico em Aristóteles: significado y fundamentación	Sentido Filosófico	Franciscanum; 56(162); 51-73; 2014-07
Framework baseado em MDA y ontologias para la representación y validación de modelos de componentes	Ontologia/Domínio de Gestão Empresarial	Rev cuba cienc informat; 8(2); 102-116; 2014-06
Prolegômenos a Toda a Administrolgia Possível: Administração – O Que é Isto?	Sentido Filosófico	Organ. Soc.; 21(71); 695-712; 2014-12
A didática como diferenciador das diferenças pelas potências dos possíveis	Sentido Filosófico	Pro-Posições; 26(1); 123-137; 2015-04
Dos Antagonismos na Apropriação Capitalista da água à sua Concepção como Bem Comum	Sentido Filosófico	Organ. Soc.; 22(73); 237-250; 2015-06
O pensamento como desordem: repercussões do legado foucaultiano	Sentido Filosófico	Pro-Posições; 25(2); 83-101; 2014-08

Cognitive models of executive functions development: methodological limitations and theoretical limitation and theoretical challenges	Propostas para o Desenvolvimento de uma Ontologia	Anal. Psicol.; 30(1); 329-336; 2014-01
Heidegger y la noción precencia (Vorwissenschaft) como propedêutica existencial. (Parte 1: El afecto del tempo)	Sentido Filosófico	Estud.filos; (50); 155-169; 2014-12
Antonio Banfi e le filosofie dela Germania del novecento	Sentido Filosófico	Trans/Form/Ação; 37(3); 201-216; 2014-12
El problema del objetivismo em la producción de conocimiento sociológico: la correspondência entre Alfred Schutz, Talcott Parsons y Harold Garfinkel	Sentido Filosófico	Cinta moebio; (51); 171-191; 2014-12
O limite da normatividade na teoria política de Ernesto Laclau	Sentido Filosófico	Lua Nova; (91); 135-167; 2014-04
Estados de cosas negativos: La perspectiva fenomenológica de Adolf Reinach	Sentido Filosófico	Veritas; (30); 65-81; 2014-03
De la evaluación instrumental a la evaluación desde la alteridade	Sentido Filosófico	Saber; 26(3); 335-342; 2014-09
Sedeas: Prototipo de Sistema expert para el diagnóstico de enfermedades autoinmunes de órgano basado em internist	Ontologia/Domínio de Doenças Auto-imunes	Rev Univ. salud.; 16(2); 207-218; 2014-12
La Estructura de las Revoluciones Científicas: cincuenta años	Sentido Filosófico	Rev. iberoam. cienc. tecnol. soc.; 8(22); 133-151; 2014-01
Linguagem e comunicação em Heidegger	Sentido Filosófico	Galáxia (São Paulo); 14(27); 85-94; 2014-06
Nacionalismo y violencia: uma explicación mecanística. Com especial referencia a las teorías de Charles Tilly y Michael Mann	Sentido Filosófico	Cultura representaciones soc; 9(18); 26-62; 2015
Interações de Rudolf Steiner com uma Educação Anticolonial	Sentido Filosófico	Educ. rev.; (56); 85-100; 2015-06
O sábio-aprendiz e o efêmero lugar da escrita: para uma ética da inventividade acadêmica	Sentido Filosófico	Educ. Pesqui.; 40(2); 483-498; 2014-06
Social Justice and the "Green" City	Sentido Filosófico	Rev. Bras. Gest. Urbana; 6(2); 158-169; 2014-08

Fonte: Própria da pesquisa

Ao se fazer a busca de publicações do período de 2014 a 2015, refinada com o tópico Ontologia obteve-se 48 ocorrências, conforme o quadro 6, dos 48 artigos disponíveis, 8 são sobre Ontologias nos seguintes domínios: Enfermagem, Jurídico, Referência de Modelo de Processos, Anjos, Causas de Morte, Genética, Gestão empresarial e Doenças Auto-imunes. Apresenta 3 artigos sobre desenvolvimento e implementação de Onto-

logias e 1 sobre uma Ontologia de Representação (Meta-ontologia) no domínio Geoespacial. Os Outros 36 artigos encontrados apresentam o termo Ontologia no seu sentido filosófico.

A BDTD (Biblioteca Digital de Teses e Dissertações) é um repositório de teses e dissertações, que partiu de uma iniciativa do IBICT (Instituto Brasileiro de Informação em Ciência e Tecnologia), vinculado ao MCT (Ministério de Ciência e Tecnologia). A BDTD estimula o registro e a publicação de teses e dissertações em meio eletrônico, desta forma dando mais visibilidade à produção científica nacional. O termo Ontologia ocorreu 1.220 vezes, apresentando 819 teses e 401 dissertações, nos idiomas português, espanhol e inglês.

Quadro 7- Análise da base de dados BDTD

TÍTULO	CARACTERÍSTICA	GÊNERO	INSTITUIÇÃO
Estudo sobre o crátilo de Platão	Sentido Filosófico	Dissertação	UFSC
OWLSUMBRP: Um método para sumarização de ontologias	Metodologia para sumarização de Ontologias	Dissertação	UFPE
MAS ONTOLOGY: uma ontologia de métodos orientados a agentes	Ontologia/Domínio de Metodologias de Sistemas Multiagentes	Dissertação	UERJ
O Solipsismo de Evaldo Coutinho	Sentido Filosófico	Dissertação	UFPE
Consciência e nadificação	Sentido Filosófico	Dissertação	UFSC
Adelmo Genro Filho e a Teoria do Jornalismo no Brasil	Sentido Filosófico	Tese	UFSC
A tese da subdeterminação da metafísica pela física	Sentido Filosófico	Dissertação	UFSC
Generidade e politicidade no último lukács /	Sentido Filosófico	Tese	UNESP
Em busca de uma existentiellvidenskab: kierkegaard e a ontologia do inter-esse	Sentido Filosófico	Tese	UNISINOS
Estudo do processo de apropriação da ontologia pela Ciência da Informação no Brasil	Produção de Ontologias na Ciência da Informação	Dissertação	UFPE
Método de argumentação para resolução colaborativa de divergências na combinação de ontologias individuais	Desenvolvimento de Ontologias	Tese	UTFPR
Uma proposta de ontologia para residências inteligentes buscando a integração de dispositivos	Ontologia/ Domínio Domiciliar	Dissertação	UTFPR
Preconceito e diálogo	Sentido Filosófico	Dissertação	UFSC

Vir-a-ser da sensibilidade	Sentido Filosófico	Dissertação	UFSC
A noção de 'realidade' no tractatus de wittgenstein	Sentido Filosófico	Dissertação	UNICAMP
Corpo, natureza, carne	Sentido Filosófico	Tese	UFSC
Proposta de ontologia para integração de sistemas de manutenção inteligentes e cadeias de suprimento de peças de reposição	Desenvolvimento de Ontologia	Dissertação	UFRGS
Anotação semântica baseada em ontologia: um estudo do português brasileiro em documentos históricos do final do século XIX	Ontologia/Domínio da Instrumentolingüística	Dissertação	UFSCAR
Folkconcept: método de suporte à modelagem conceitual de ontologias a partir da aquisição de conhecimentos de folksonomias	Desenvolvimento de Ontologia	Tese	UTFPR
Ontologia aplicada no mapeamento de classe de dados para a melhoria do intercâmbio de dados na cotonicultura no Brasil.	Ontologia/Domínio da Cotonicultura	Tese	USP
Modelagem semântica de processos industriais com aplicações	Ontologia/Domínio da Automação industrial	Tese	UFRN
A sociedade é inefável : sobre a individualidade do protagonista do romance "os anos de aprendizado de wilhelm meister" (1795/96), de goethe	Sentido Filosófico	Tese	UNICAMP
As bases ontológicas do processo de apropriação do conhecimento e seus desdobramentos para a educação infantil	Sentido Filosófico	Tese	UFSC
Ação e artifício	Sentido Filosófico	Tese	UFSC
Aspectos formais e ontológicos da filosofia da ciência de Aristóteles	Sentido Filosófico	Dissertação	UNICAMP
Subjetividade e finitude em ser e tempo	Sentido Filosófico	Dissertação	UFG
Uma abordagem ontológica para recomendação de equipes qualificadas em projetos de software	Desenvolvimento de Ontologia	Dissertação	UFPB
Temporalidade e facticidade: a vida cristã primitiva no jovem heidegger	Sentido Filosófico	Dissertação	UFPB
Realismo e racionalidade: o otimismo epistêmico em questão	Sentido Filosófico	Tese	USP
Agostinho e os maniqueus: análise a partir 'das duas almas'	Sentido Filosófico	Dissertação	USP
João Gilberto Noll e a estética do não-eu	Sentido Filosófico	Tese	UNICAMP
Configurações da política de integração	Sentido Filosófico	Tese	UFSC

A mediação como fundamento da educação física escolar: possibilidades e limitações para a práxis educativa /	Sentido Filosófico	Dissertação	UNESP
Do realismo burguês ao realismo socialista: um estudo sobre a questão da herança cultural no pensamento de Lukács nos anos 1930	Sentido Filosófico	Tese	USP
A alienação em Lukács: fundamentos para o entendimento do complexo da educação.	Sentido Filosófico	Tese	UFC
Teorias da conduta: uma análise crítica	Sentido Filosófico	Dissertação	UERJ
A gênese das categorias trabalho e habitus : pistas ontológicas apreendidas do diálogo entre os "jovens" Lukács e Bourdieu	Sentido Filosófico	Dissertação	UNICAMP
Geografia e individuação	Sentido Filosófico	Dissertação	USP
Um Samuel Rawet qualquer: pensamento-prosa	Sentido Filosófico	Dissertação	USP
Ubigroup: um modelo de recomendação ubíqua de conteúdo para grupos de aprendizes	Desenvolvimento de Ontologia	Dissertação	UNISINOS
A filosofia da alteridade ou o problema da ausência na filosofia contemporânea	Sentido Filosófico	Dissertação	UFU
Descoberta dinâmica, sensível ao contexto, de serviços web	Desenvolvimento de Ontologias	Dissertação	UFSCAR
Da percepção à visão radical do mundo: a condição de abertura do plano de imagens no percurso de merlau-ponty	Sentido Filosófico	Tese	USP
Marcação de regiões de interesse em 3d sobre imagens radiológicas utilizando a web	Ontologia/Domínio de Radiologia	Dissertação	USP

Fonte:Própria da pesquisa

Ao se fazer a busca de publicações do período de 2014 a 2015, refinaada com o tópico Ontologia, obteve-se 44 ocorrências, conforme o quadro 7, dos 44 documentos disponíveis (17 teses e 27 dissertações), 6 são sobre Ontologias nos seguintes domínios: Metodologia de Sistemas Multiagentes, Domiciliar, Instrumentolinguística, Cotonicultura (produção de algodão), Automação industrial e Radiologia. Apresenta 6 documentos sobre desenvolvimento e implementação de Ontologias, 1 sobre Metodologia de Sumarização de Ontologias, 1 sobre a Produção de Ontologias na área de Ciência da Informação. Os Outros 30 documentos encontrados apresentam o termo Ontologia no seu sentido filosófico.

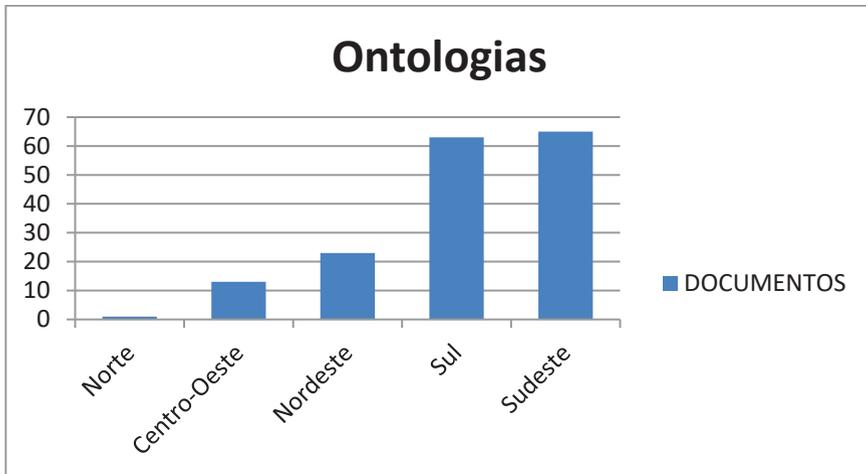
Quadro 8 - Resultado das buscas

CATEGORIAS BASES	Documentos recuperados	Gênero	Resultados Duplicados	Documentos relevantes
CAPES	18	Artigo	-	4
LILACS	12	Artigo	-	8
SCIELO	48	Artigo	1	12
BDTD	44	Teses e Dissertações	-	14

Fonte: Própria pesquisa

A análise revela que a produção de Ontologias tem crescido nos últimos anos e a criação de domínios nas mais variadas áreas tem se desenvolvido e se mostrado presente em muitas pesquisas. Ao analisar especificamente a base de dados da BDTD, pode-se observar que, ainda há uma grande diferença em relação à produção científica e desenvolvimento de Ontologias entre as regiões Norte e Nordeste e as demais, a grande concentração deste tipo de produção está nas regiões Sul e Sudeste.

Gráfico 1 - Produção e Desenvolvimento de Ontologias por região



Fonte: Própria pesquisa

Ao fazermos um recorte de produção disponibilizada na BDTD, entre 2014 e 2015, sem refinar termos, podemos observar, conforme o quadro 9 que ainda assim é grande a diferença de produção por região. De um total de 165 documentos, 65 são produções da região Sudeste, 63 da região Sul, 23 da região Nordeste, 13 da região Centro-Oeste e apenas 1 da região Norte.

Quadro 9 - Publicações de Ontologia por região

REGIÃO	DOCUMENTOS	SENTIDO FILOSÓFICO	RELEVANTES SOBRE O TEMA
Sudeste	65	44	21
Sul	63	30	33
Nordeste	23	11	12
Centro-Oeste	13	9	4
Norte	1	-	1

Fonte: Própria pesquisa

A análise da produção científica sobre Ontologias na área da Ciência da Informação, representada por meio das publicações disponíveis na BDTD, permite identificar que no período de 2014 a 2015, a região Nordeste tem mantido certo equilíbrio em relação às publicações com o termo Ontologia em seu sentido filosófico, uma vez que, 52% dos documentos disponíveis para pesquisa são sobre o uso, metodologia e desenvolvimento de Ontologias.

O USO DE ONTOLOGIAS NA REPRESENTAÇÃO DO CONHECIMENTO EM BIBLIOTECAS UNIVERSITÁRIAS DAS REGIÕES NORTE E NORDESTE

Segundo o IBGE (Instituto Brasileiro de Geografia e Estatística) a região Norte³ é a maior das cinco regiões do Brasil, com uma extensão territorial de 3 853 676, 948 km². É composta por sete estados: Acre, Amapá, Amazonas, Pará, Rondônia, Roraima e Tocantins. Já a região Nordeste⁴ tem a segunda maior população do Brasil e o terceiro maior território com uma extensão territorial de 1 554 291,744 km². Possui nove estados: Alagoas, Bahia, Ceará, Maranhão, Paraíba, Piauí, Pernambuco, Rio Grande do Norte e Sergipe (figura 9).

³ https://pt.wikipedia.org/wiki/Regi%C3%A3o_Norte_do_Brasil

⁴ https://pt.wikipedia.org/wiki/Regi%C3%A3o_Nordeste_do_Brasil

Figura 9: Mapa da região Norte e Nordeste



Fonte: http://www.integralagro.com.br/_site/atuacao/index.htm

As Bibliotecas das Universidades Federais presentes em cada estado das regiões Norte e Nordeste, conforme quadro 9, constituem o universo da pesquisa, tendo como critério de seleção as primeiras universidades federais das capitais. Dentre as bibliotecas presentes nas 7 Universidades Federais da região Norte, apenas 3 constitui a amostra: Sistema de Bibliotecas da Universidade Federal do Amazonas – UFAM, Biblioteca Central Prof^a Maria Auxiliadora de Souza Melo – UFRR e Biblioteca Universitária da Universidade Federal do Pará – UFPA.

Quadro 10 - Universidades Federais das regiões Norte e Nordeste

Região Norte	Região Nordeste
Universidade Federal do Acre – UFAC	Universidade Federal de Alagoas – UFAL
Universidade Federal do Amapá – UNIFAP	Universidade Federal da Bahia – UFBA
Universidade Federal do Amazonas – UFAM	Universidade Federal do Ceará – UFC
Universidade Federal do Pará – UFPA	Universidade Federal do Maranhão – UFMA
Universidade Federal de Rondônia – UNIR	Universidade Federal da Paraíba – UFPB
Universidade Federal de Roraima – UFRR	Universidade Federal de Pernambuco – UFPE
Universidade Federal de Tocantins – UFT	Universidade Federal do Piauí – UFPI
	Universidade Federal do Rio Grande do Norte – UFRN
	Universidade Federal de Sergipe – UFS

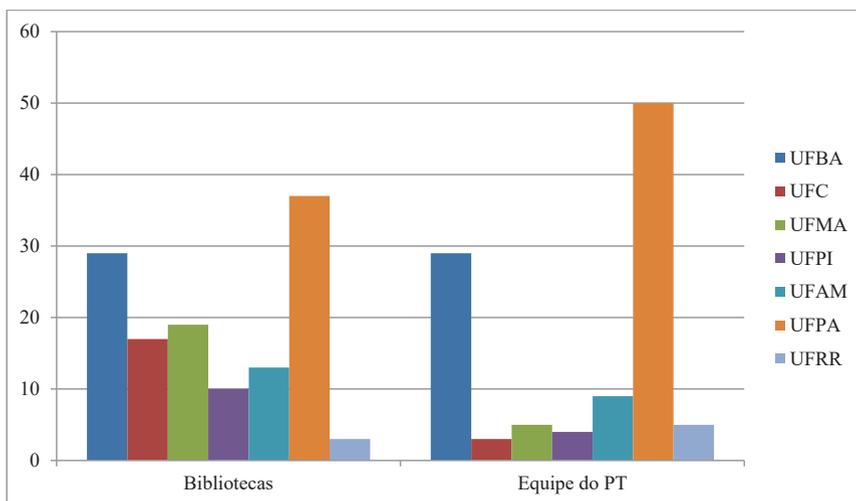
Fonte: Própria da pesquisa

Em relação à região Nordeste, das bibliotecas presentes nas 9 Universidades Federais da região, apenas quatro bibliotecas constitui a amostra: Biblioteca da Faculdade de Arquitetura- UFBA, Biblioteca Universitária da Universidade Federal do Ceará – UFC, Sistema de Bibliotecas da Universidade Federal do Maranhão – UFMA, Biblioteca Universitária da Universidade Federal do Piauí – UFPI.

As bibliotecas universitárias tiveram que acompanhar os avanços da Tecnologia da Informação e adequar seus espaços informacionais em favor da qualidade no atendimento aos usuários. Tem a missão de dar suporte às atividades de ensino, pesquisa ou extensão e para garantir que tal suporte seja bem efetuado, deve ter uma equipe de bibliotecários que além de ter a consciência de seu papel, assegure que a informação chegue a tempo hábil para seus usuários.

No contexto de Bibliotecas Universitárias, o tratamento temático da informação é realizado pela indexação e catalogação de assuntos e assim, pode-se articulá-las como sendo a mesma atividade, pois os resultados de seus processos resultaram na análise de assunto dos documentos. (REDIGOLO, 2010, p.37).

Gráfico 2 – Quantidade de bibliotecas e equipe do Processamento Técnico



Fonte: Própria da pesquisa

O Sistema de Bibliotecas da Universidade Federal do Amazonas possui um total de 13 bibliotecas, no entanto não há setor de processamento técnico em todas elas, apenas em 4 localizadas no Campus e em 5 localizadas no campi no interior, que conta com 1 bibliotecário em cada uma

delas. Entre os instrumentos utilizados pela equipe estão CDU, AACR2, CUTTER, Tesouros, Catálogos da Rede *Pergamum*, da Biblioteca Nacional (BN) e da *Library of Congress* (LC), além dos cabeçalhos de assunto e de autoridades da BN e LC, entre outros específicos como BVS e EMBRAPA.

O Sistema de Bibliotecas da Universidade Federal do Pará é composto por 37 bibliotecas e apesar de não possuir o setor de processamento técnico em todas elas, possui uma equipe de 50 bibliotecários na equipe responsável pelo processamento técnico, que utilizam CDU, AACR2 e CUTTER.

As Bibliotecas da Universidade Federal de Roraima são um total de 3, não há setor de processamento técnico em todas elas, sendo restrito apenas à Biblioteca Central Prof^a Maria Auxiliadora de Souza Melo que possui na equipe 1 bibliotecária, 1 técnica e 1 bolsista. Utilizam CDU, AACR2, CUTTER e bases de dados de outras IFES: UFSC, UNB, Rede bibliodata e UFV.

O Sistema de Bibliotecas da UFBA é composto por 29 bibliotecas e em todas elas há o setor de processamento técnico, no entanto a quantidade de bibliotecários da equipe varia de acordo com o tamanho da biblioteca e utilizam CDU, CDD, AACR2 e CUTTER.

O Sistema de Bibliotecas da UFC é composto por 17 bibliotecas, em todas há o setor de processamento técnico, que conta com uma equipe de 3 bibliotecários, que utilizam CDD, AACR2 e tabela PHA.

O Sistema de Bibliotecas da UFMA é composto por 19 bibliotecas e não há o setor de processamento técnico em todas elas, uma vez que o processo é centralizado, sendo realizado em setor específico da Biblioteca Central, que é formado por uma equipe de 5 bibliotecários. Entre os instrumentos utilizados estão CDU, AACR2, CUTTER, Tesouros e lista de cabeçalho de assuntos de outras instituições.

O Sistema de Bibliotecas da UFPI é formado por 10 bibliotecas e não há o setor de processamento técnico em todas as unidades, a equipe de processamento técnico é formada por 4 bibliotecários e se concentra na biblioteca central, utilizam CDD, AACR2 e CUTTER.

Essas atividades são inerentes ao Setor de Processamento Técnico da Unidade de informação, responsável pelos processos de catalogação e indexação, e compreendem o uso de instrumentos da área de Biblioteconomia tais como: Classificação Decimal de Dewey – CDD, Código de Catalogação Anglo-Americano, 2.ed. – AACR2, Tabela Cutter-Sanborn, Tesouro – Cabeçalhos Controlados de Autoridade e de Assunto, bem como ferramentas de Tecnologias de Comunicação e Informação para automatização do acervo. (PEREIRA; LAURINDO; SANTIAGO, 2011).

Quadro 11 – Caracterização das Bibliotecas

Instituição	Bibliotecas	Bibliotecários do PT	Instrumentos Utilizados
UFAM	13	9	CDU, AACR2, CUTTER, Tesouros e Catálogos
UFPA	37	50	CDU, AACR2 E CUTTER
UFRR	3	1	CDU, AACR2, CUTTER e Bases de dados
UFBA	29	29	CDU, CDD, AACR2 e CUTTER
UFC	17	3	CDD, AACR2 e tabela PHA
UFMA	19	5	CDU, AACR2, CUTTER, Tesouros e lista de cabeçalho de assunto
UFPI	10	4	CDD, AACR2 e CUTTER

Fonte: Própria da pesquisa

O bibliotecário que atua no setor de processamento técnico é o responsável por realizar o tratamento do documento, extraindo do mesmo, os termos que melhor irão representá-lo, a fim de garantir um alto índice de recuperação da informação solicitada pelo usuário. Desta forma, diante dos avanços tecnológicos, da explosão informacional e da variedade de mídias e suportes, deve buscar acompanhar as mudanças, para poder assegurar seu papel de mediador da informação.

Segundo Anna (2013, p. 8), “[...] o tratamento da informação em espaço digital exigirá o conhecimento do uso das novas formas de representação, uma vez que os objetos digitais, diferentemente dos físicos, apresentam diferenças de formas e estrutura o que irá requerer um domínio específico para seu gerenciamento”.

Atualmente, no contexto da *web* e das bibliotecas digitais, surge a necessidade de se buscar outros instrumentos de representação descritiva, estudos apontam para o uso de metadados, que além de possibilitarem a padronização das formas de representação do conteúdo informacional, objetivam garantir a interoperabilidade entre sistemas.

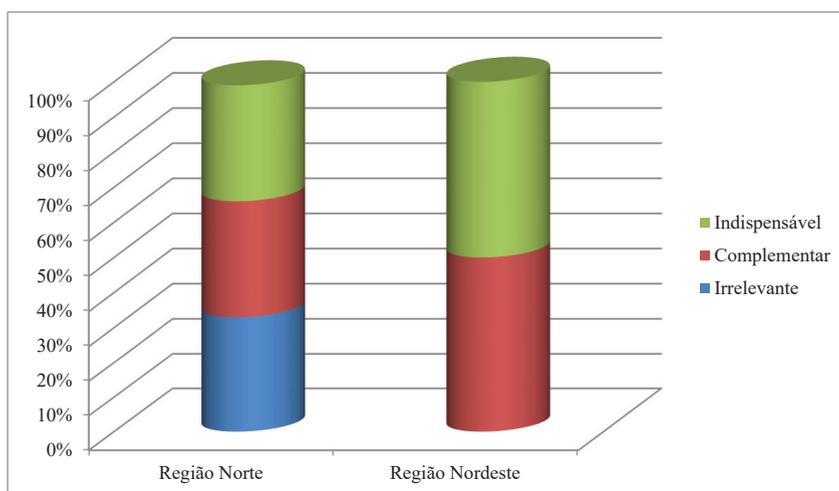
O uso de FRBR (Requisitos Funcionais para Registros Bibliográficos), por exemplo, tem crescido nos últimos anos, uma vez que, o seu foco no processo de catalogação está no usuário e não no conteúdo, além de disponibilizar uma tecnologia clara e inovadora. Pacheco (2014), afirma que “[...] o modelo conceitual FRBR, não é um código de catalogação, nem um formato para registros bibliográficos, mas sim um modelo entidade-relacionamento que propõe os requisitos para um registro bibliográfico e os relacionamentos entre suas entidades”. O uso de FRBR tem facilitado o acesso do usuário à informação, uma vez que, trabalha com entidades,

relacionamentos e atributos que se relacionam com os catálogos informatizados.

Sobre o uso de Ontologias na organização e representação da informação

As Ontologias possibilitam a comunicação entre pessoas em relação a determinado conhecimento, visto que, permitem raciocínio e entendimento sobre um domínio. Segundo Moraes e Ambrósio (2007, p.5), “As ontologias formam um vocabulário de consenso que permite representar conhecimento de um domínio em seu nível mais alto de abstração, possuindo, desta forma, potencial de reutilização”.

Gráfico 3 – Utilização de Ontologias no Processamento Técnico



Fonte: Própria pesquisa

No âmbito das bibliotecas universitárias, o bibliotecário deve buscar constante atualização e familiarização com os novos instrumentos e ferramentas advindas dos avanços tecnológicos, uma vez que, lidam de forma direta com diversos tipos de usuários, desde universitários a professores e pesquisadores, das mais variadas áreas. E, portanto, o conhecimento e uso de domínios construídos em áreas específicas do conhecimento permitirão que os documentos sejam representados com os termos adequados, contribuindo, assim, para a integração bibliotecário – usuário.

Ao compararmos a Ontologia aos instrumentos tradicionais de catalogação, podemos observar a amplitude das Ontologias, pois não se

limitam aos relacionamentos hierárquicos, uma vez que presam pelo relacionamento associativo entre termos variados.

Quadro 12 – Uso de Ontologias

Instituição	Primeiro Contato	Uso de Ontologias	Domínios conhecidos
UFAM	Biblioteca	Indispensável	Meio Ambiente e Agricultura
UFPA	Encontros e Congressos	Complementar	-
UFRR	Curso de Biblioteconomia	Irrelevante	-
UFBA	Encontros e Congressos	Complementar	-
UFC	Encontros e Congressos	Complementar	-
UFMA	Encontros e Congressos	Indispensável	-
UFPI	Curso de Biblioteconomia	Indispensável	Direito

Fonte: Própria da pesquisa

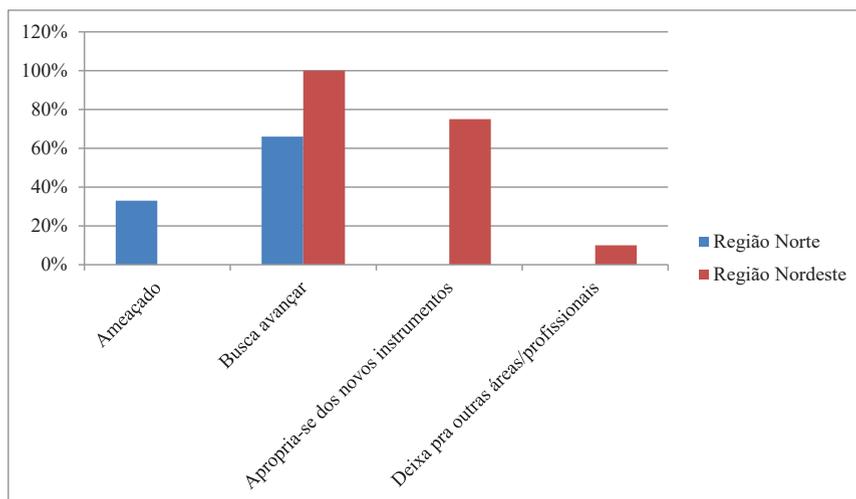
Como podemos observar o grau de familiarização com Ontologias pelos bibliotecários das universidades da amostra, ainda é bem pequeno, acredito que o não uso de Ontologias nos processos de representação da informação, é devido à falta de conhecimento sobre esta ferramenta, uma vez que o contato ainda é reduzido. Cabe ressaltar que não se trata de falta de infra-estrutura das universidades de modo geral, visto que outras áreas como a Ciência da Computação tem avançado e produzido muito sobre esta temática.

Surge, então, a necessidade de, enquanto profissionais da informação, buscar acompanhar as inovações tecnológicas, no que diz respeito, ao tratamento da informação em outros formatos, além do impresso. A maioria dos bibliotecários parou no tempo, agarrando-se a teorias e práticas de anos atrás, há a necessidade de busca e apropriação de novos saberes nesta área, pois nós dominamos os aspectos de extração da informação documental, no entanto, devemos manter a educação continuada a fim de poder ter subsídios para representar e organizar a informação em outros suportes.

Podemos observar que a relação dos bibliotecários com este instrumento de organização da informação denominado Ontologia, ainda não é dos mais familiares, talvez por ser mais trabalhado por outras áreas do

conhecimento como a Ciência da Computação e a Inteligência Artificial, no entanto, devido às mudanças nos suportes informacionais e na grande variedade de documentos disponibilizados na internet, seja necessário um novo olhar dessa classe de profissionais e uma apropriação das novas ferramentas indispensáveis no momento da organização e representação da informação e do conhecimento.

Gráfico 4 – Postura do bibliotecário no novo cenário informacional



Fonte: Própria pesquisa

A quantidade de documentos disponibilizados na *web* é enorme e a internet tem sido uma grande fonte de busca de informação. O bibliotecário deve se especializar a fim de contribuir com a organização da informação na *web*. Na região Norte os bibliotecários do Sistema de Bibliotecas da UFAM contribuem com esta organização utilizando instrumentos que melhor representem a informação na *web*, trocando informações com os pares e buscando colaboração com profissionais de outras áreas com a Ciência da Computação e Inteligência Artificial, sempre buscando acompanhar os avanços para melhor executar suas atividades profissionais.

O Sistema de Bibliotecas da UFPA tem contribuído para a organização da informação na *web* por meio da utilização de instrumentos que melhor representem este tipo de informação, segundo os dados da pesquisa, o bibliotecário sente ameaçado pelo avanço das novas tecnologias e quando indagado sobre a representação do volume de informações na *web*, dá a seguinte sugestão: “Utilizando folksonomias na representação temática da informação”.

Em relação à forma de contribuição na organização da informação na web por parte dos bibliotecários do Sistema de Bibliotecas da UFRR, afirmam buscar colaboração com profissionais de outras áreas e buscam acompanhar os avanços para melhor executar suas atividades profissionais.

Para os bibliotecários da região Nordeste, especificamente os do Sistema de Bibliotecas da UFBA apropriam-se do conhecimento necessário por meio da educação continuada e troca de informação com os pares. E diante desse processo de representação da informação na web buscam acompanhar os avanços e apropriar-se de instrumentos que facilitem o mesmo.

Os bibliotecários do Sistema de Bibliotecas da UFC além de utilizarem instrumentos que melhor representem este tipo de informação, estão sempre trocando informação com os pares e colaboração com profissionais de outras áreas e diante desse processo buscam acompanhar os avanços da área e a apropriação dos instrumentos capazes de facilitar a representação da informação na web. Quanto à organização da informação na web, fazem as seguintes sugestões: “Estudo dos usuários; Política de indexação; Vocabulário controlado nas diversas áreas do conhecimento, a fim de que os usuários recuperem as suas necessidades informacionais; E busca pelo desenvolvimento e aprimoramento de ferramentas e metodologias que proporcionem uma recuperação eficaz”.

O Sistema de Bibliotecas da UFMA tem contribuído na organização da informação na web por meio da educação continuada, uso de instrumentos que melhor representem este tipo de informação, troca de informação com os pares e com a colaboração com outros profissionais. E se vê nesse processo buscando acompanhar os avanços e a apropriação de instrumentos capazes de facilitar a representação da informação na web. E para que este processo seja feito de forma satisfatória dão a seguinte sugestão: “O uso de buscadores ou meta buscadores baseados em ontologias possibilitaria a precisão na recuperação da informação pelos usuários”.

E por fim os bibliotecários do Sistema de Bibliotecas da UFPI têm contribuído com o processo de organização da informação na web por meio da colaboração com profissionais da área da Ciência da Computação e Inteligência Artificial e diante deste processo buscam acompanhar os avanços e quanto à organização da informação na web afirmam que: “É importante que a informação disseminada na internet seja indexada. O Google deu um grande salto de qualidade na contratação de profissional da informação para indexar seus conteúdos”.

Quadro 13 – Relação do bibliotecário com Ontologias

Instituição	Como a biblioteca contribui com a organização da informação na web	Como o bibliotecário se vê nesse processo de organização
UFAM	Utilizando instrumentos adequados, trocando informações com os pares e com a colaboração de profissionais de outras áreas	Busca acompanhar os avanços das novas tecnologias
UFPA	Utilizando instrumentos adequados	Sente-se ameaçado pelo avanço das novas tecnologias
UFRR	Buscando colaboração com profissionais de outras áreas	Busca acompanhar os avanços das novas tecnologias
UFBA	Trocando informações com os pares e por meio da educação continuada	Busca acompanhar os avanços e se apropriar dos instrumentos que facilitem a representação na web
UFC	Utilizando instrumentos adequados, trocando informações com os pares e com a colaboração de profissionais de outras áreas	Busca acompanhar os avanços e se apropriar dos instrumentos que facilitem a representação na web
UFMA	Utilizando instrumentos adequados, trocando informações com os pares, colaboração de profissionais de outras áreas e educação continuada	Busca acompanhar os avanços e se apropriar dos instrumentos que facilitem a representação na web
UFPI	Buscando colaboração com profissionais de outras áreas	Busca acompanhar os avanços das novas tecnologias

Fonte: Própria da pesquisa

A pesquisa apresenta entre os dados que, os bibliotecários estão buscando acompanhar os avanços a fim de melhor executar suas atividades, por meio da educação continuada e da apropriação de instrumentos que facilitem a representação na *web*. No entanto, podemos perceber que se trata apenas de um discurso, pois as evidências da pesquisa apontam que os bibliotecários ainda não se apropriaram das novas ferramentas de representação da informação, afirmam não ter um conhecimento sobre domínios de Ontologias, que é bastante intrigante, visto que estão no ambiente da universidade e atendem a estudantes e pesquisadores de áreas variadas e para que o atendimento seja satisfatório deve haver esse conhecimento para poder fornecer a informação necessária.

Existem grupos de pesquisa de catalogação, encontros como o ENACAT (Encontro Nacional de Catalogadores), EEPC (Encontro de Estudos e Pesquisas em Catalogação), e inúmeros congressos e eventos que abordam esta temática sobre os novos instrumentos de representação da informação. Os bibliotecários devem manter constante sua atualização nesta área, que por sinal é sua principal característica, caso contrário, será substituído por profissionais de outras áreas, que nem tem tanta proprie-

dade na análise documental, mas que busca acompanhar os avanços tecnológicos.

O bibliotecário não deve sentir-se ameaçado diante das novas tecnologias ou ter medo de perder seu espaço, no entanto, deve buscar enxergar as oportunidades que estão surgindo neste novo cenário informacional. Deve ser um profissional dinâmico e inovador, sem limitar-se às mesmas atividades, uma das características que devem estar presentes no perfil profissional do bibliotecário é a busca constante por atualização, uma vez que é caracterizado como um profissional da informação, portanto deve saber manipular as tecnologias a seu favor. Também deve buscar o trabalho integrado com outras áreas a fim de compartilhar o aprendizado com outros profissionais, explorando e integrando as especialidades de cada um e soma-las ao bom uso das novas tecnologias.

CONCLUSÃO

A organização da informação tem se tornado um grande desafio nos dias atuais, diante das mais variadas mídias e suportes. Tal organização compreende um processo de representação da informação cujo objetivo primordial é a sua recuperação pelo usuário. Diante da diversidade de meios onde a informação é representada, os profissionais responsáveis por esta tarefa devem acompanhar os avanços em tecnologia para alterar os meios onde a representação é realizada.

Juntamente com a internet e as Bibliotecas Digitais surgiram novos desafios, uma vez que o novo suporte informacional gera outra forma de busca, que permite além do acesso a interação com informação procurada. Embora a representação da informação continue com a mesma finalidade de descrever e tratar documentos, a apropriação de novas ferramentas é indispensável no ambiente da *web*. A partir deste contexto, exige-se do bibliotecário uma nova postura, novos saberes, visto que é denominado um profissional da informação, e para tal, necessita conhecer os novos instrumentos de representação, a fim de organizar a informação para atender, de forma satisfatória seus usuários.

Quanto à apropriação de Ontologias pelos bibliotecários das universidades federais das regiões Norte e Nordeste do Brasil, a pesquisa revelou que, estes profissionais ainda não se apropriaram de fato, de tal instrumento de representação da informação e poucos conhecem domínios específicos e os utilizam no processo de processamento técnico nas bibliotecas em que atuam e na organização da informação em seus Repositórios Institucionais.

Ao analisar como os bibliotecários estão inseridos na organização da informação na *web*. Os bibliotecários que compõem a amostra da pesquisa afirmam que estão sempre em busca da apropriação de novos conhecimentos, por meio da educação continuada, buscam parceria com profissionais de outras áreas como Ciência da Computação e Inteligência Artificial. Além de buscarem acompanhar os avanços da tecnologia para melhor desenvolver suas atividades profissionais. No entanto, a própria pesquisa aponta que a apropriação das novas ferramentas e instrumentos de representação da informação, ainda está longe de ser algo presente no perfil profissional de muitos bibliotecários, que afirmam não usar Ontologias na organização da informação e até mesmo desconhecer a gama do que tem sido produzido nos mais variados domínios do conhecimento.

Em se tratando das ferramentas tecnológicas que contribuem para a organização da informação na *web* tais como taxonomias e *folksonomia*, *frames* semânticos e Ontologias, traz as principais características de tais ferramentas dentro da Web Semântica, com o destaque para as Ontologias, retratadas como a linguagem da Web Semântica, especificando sua estrutura em camadas. E sobre as técnicas e tecnologias para o desenvolvimento de Ontologias, apresenta o seu processo de criação, suas formas de representação, grau de formalismo e tipos. Apresenta ainda, algumas metodologias de desenvolvimento, a pesquisa revela que a metodologia mais utilizada no desenvolvimento de Ontologias é a *Methontology* (cuja construção é baseada a partir do conhecimento de um domínio) juntamente com o editor Protegé.

Em relação aos domínios de Ontologias disponíveis, foram apresentados alguns domínios do conhecimento que tem Ontologias, como o domínio Jurídico, domínio em Ciências Biológicas, domínio em Nanociência e Nanotecnologia, domínio de Futebol e o domínio do Patrimônio Cultural de São Luís, identificando que existem muitas áreas, que tem produzido Ontologias para facilitar a representação do conhecimento de um domínio específico. O que é um dado conflitante com a análise de um dos aspectos do instrumento utilizado na pesquisa, visto que, a maioria dos bibliotecários que constituem a amostra, afirma não conhecer um domínio específico de Ontologias, o que permite concluirmos que profissionais de outras áreas como a Ciência da Computação e a Inteligência Artificial, têm avançado nesta produção e busca pelo uso e desenvolvimento de Ontologias.

Sobre os domínios de Ontologias disponíveis e a produção na área da Ciência da Informação, foi mapeado a partir da análise de sua produção e representação nas bases dados: CAPES, LILACS, SCIELO E BDTD, que

mostrou que a quantidade de publicações sobre Ontologias tem crescido nos últimos anos e a criação de domínios nas mais variadas áreas tem se desenvolvido e se mostrado presente em muitas pesquisas. Ao analisar especificamente a base de dados da BDTD pode-se observar que ainda há uma grande diferença em relação à produção científica e desenvolvimento de Ontologias entre as regiões Norte e Nordeste e as demais.

A pesquisa revela que da amostra de bibliotecas das duas regiões nenhuma produziu algum tipo de Ontologia, poucas utilizam no processamento técnico e a maioria afirma não conhecer domínio que utilize Ontologias. No entanto, todas as bibliotecas afirmam buscar acompanhar os avanços na área, a fim de contribuir com a organização da informação na *web* e desconhecem a produção disponibilizada em suas próprias bases de dados.

REFERÊNCIAS

ALBUQUERQUE, A.; MESQUITA, D.; COSTA, L. Folksonomia: uma nova modalidade de indexação e recuperação da informação na web. In: Encontro Nacional de Estudantes de Biblioteconomia, Documentação, Gestão e Ciência da Informação, 33., 2010, João Pessoa. **Anais...** João Pessoa: ENEBD, 2010.

ALMEIDA, Maurício B.; BAX, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ci. Inf.**, Brasília, v. 32, n. 3, p. 7-20, set./dez. 2003.

ANJOS, Liane dos. **Sistemas de classificação do conhecimento na Filosofia e na Biblioteconomia: uma visão histórico-conceitual crítica com enfoque nos conceitos de classe, de categoria e de faceta.** 291f. 2008. Tese (Doutorado em Ciência da Informação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2008.

ANTONIOU, Grigoris; GROTH, Paul; HARMELEN, Frank Van; HOEKSTRA, Rinke. **A semantic web primer.** Massachusetts: The Mit Press, 2012.

BARBOSA, Alice. **Teoria e prática dos sistemas de classificação bibliográfica.** Rio de Janeiro: Instituto Brasileiro de Bibliografia e Documentação, 1969.

BURKE, Peter. A classificação do conhecimento: currículos, bibliotecas e enciclopédias. In: **Uma história social do conhecimento: de Gutenberg a Diderot.** Rio de Janeiro: Jorge Zahar, 2003. p. 78-108.

CARVALHEIRA, Luiz Carlos da Cruz. **Método semi-automático de construção de ontologias parciais de domínio com base em textos**. 143f. 2007. Dissertação (Mestrado em Engenharia) - Universidade de São Paulo, São Paulo, 2007.

CERVANTES, B. M. N. **Terminologia no processo de inteligência competitiva**: estudo teórico e metodológico. Londrina: EDUEL, 2006.

CONCEIÇÃO, Valdirene Pereira da. **Modelagem léxico-ontológica do domínio "Patrimônio cultural de São Luís do Maranhão"**. 191f. 2011. Tese (Doutorado em Linguística e Língua Portuguesa) - Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraguara, 2011.

DZIEKANIAK, Gisele. A organização da informação e comunicação científica: implicações para os profissionais e usuários da informação. **Em questão**, Porto Alegre, v. 16, n. 1, p. 45-59, jan./jun. 2010.

FARIA, Mário Joaquim Firmino Leite. **Definição de uma ontologia aplicada ao futebol**. 165f. 2009. Dissertação (Mestrado Integrado em Engenharia Electrotécnica e de Computadores Major Automação) - Faculdade de Engenharia, Universidade do Porto, Porto, 2009.

FEITOSA, Ailton. **Organização da informação na web**: das tags à web semântica. Brasília: Thesaurus, 2006.

FREITAS, Larissa Astrogildo de; VIEIRA, Renata. Ontologias e língua portuguesa. In: CELSUL, 8., 2008, Porto Alegre. **Anais...** Porto Alegre: CELSUL, 2008.

FURGERI, Sérgio. O papel das linguagens de marcação para a Ciência da Informação. **Transinformação**, Campinas, v. 18, n. 3, p. 225-239, set./dez. 2006.

INTEGRAL AGRO. **Mapa da região Norte e Nordeste**. 2015. Disponível em: <http://www.integralagro.com.br/site/atuacao/index.htm>. Acesso em: 25 maio 2021.

KASAMA, Deni Yuzo. **Estruturação do conhecimento e relações semânticas**: uma ontologia para o domínio da nanociência e nanotecnologia. 180f. 2009. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2009.

LADEIRA, Marcelo. **Representação de conhecimento e redes de decisão**. Porto Alegre: CPGCC da UFRGS, 1997.

LIMA, Vânia Mara Alves. A organização do conhecimento no domínio da Ciência da Informação: o mapa conceitual e terminológico como instrumento referencial para o ensino e a pesquisa. In **CID: Revista de Ciência da Informação e Documentação**, v. 4, n. 1, p. 26-43, 2013.

LIMA-MARQUES, Mamede. **Ontologias**: da filosofia à representação do conhecimento. Brasília: Thesaurus, 2006.

MÁCULAN, Benildes et al. Taxonomia, folksonomia, acessibilidade usabilidade: proposta de interseção na área de organização do conhecimento, com foco na recuperação da informação. In: Seminário em Ciência da Informação, 3., 2009, Londrina. **Anais...** Londrina: UEL, 2009.

MORAIS, E. A. M.; AMBROSIO, A. P. L. **Ontologias**: conceitos, usos, tipos, metodologias, ferramentas e linguagens. Goiânia: Universidade de Goiás, 2007. (Relatório de pesquisa).

MORON, Robson. **Ontologias artificiais no universo jurídico**. 77f. 2010. Monografia (Graduação em Direito) -Universidade Federal do Paraná, Curitiba, 2010.

OLIVEIRA, R. M.V. B. **Web semântica**: novo desafio para os profissionais da informação. Campinas: Puccamp, 2002.

PACHECO, Elisabete Moraes. **Entidades e relacionamentos do FRBR**: uma discussão a partir da obra Ulysses de James Joyce. 70f. 2014. Monografia (Graduação em Biblioteconomia) - Universidade Federal do Rio Grande, Rio Grande, 2014.

PEREIRA, Ana Maria; LAURINDO, Danielle Bianchi Rachadel; SANTIAGO, Silvana Adir. A representação descritiva e temática dos estoques informacionais da BPSC: relato de experiência. **Revista ACB**, V. 16, n.1, 2010.

PICKLER, Maria Elisa Valentim. Web semântica: ontologias como ferramentas de representação do conhecimento. **Perspect. ciênc. inf.**, v. 12, n. 1, 2007.

PIEIDADE, Maria Antonieta Requião. **Introdução à teoria da classificação**. 2. ed. Rio de Janeiro: Interciência, 1983.

RAMOS JUNIOR, Hélio Santiago; ROVER, Aires José. **Ontologias aplicadas ao direito e ao governo eletrônico**: experiências e projetos.

2009. Disponível em: <http://buscalegis.ufsc.br/revistas/index.php/observatoriodoegov/article/view/25/25>. Acesso em: 09 ago. 2021

RAUTENBERG, Sandro; GOMES FILHO, Antonio Costa; TODESCO, José Leomar; OSTUNI-GAUTHIER, Fernando Álvaro. Ferramenta ontoKEM: uma contribuição à Ciência da Informação para o desenvolvimento de ontologias. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 15, n. 1, 2010.

REDIGOLO, F. M. **O processo de análise de assunto na catalogação de documentação**: a perspectiva sociocognitiva do catalogador em contexto de Biblioteca Universitária. 2010. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

SALES, Luana Farias. **Ontologias de domínio**: estudo das relações conceituais e sua aplicação. 140f. 2006. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal Fluminense, Niterói, 2006.

SALES, R. Representação de domínios em Biblioteconomia e Arquivologia. In: SILVA, F. C.; SALES, R. (Orgs.). **Cenários da organização do conhecimento**: linguagens documentárias em cena. Brasília:Thesaurus, 2011.

SALES, Rodrigo de; CAFE, Lígia. Diferenças entre tesouros e ontologias. **Perspectivas em Ciência da Informação**, v. 14, n. 1, p. 99-116, jan./ abr. 2009.

SANTA ANNA, Jorge. A (r)evolução digital e os dilemas para a catalogação: os cibertecários em atuação. **Rev. Digit. Bibliotecon. Cienc. Inf.**, Campinas, v. 13, n. 2, maio/ago. 2015.

SILVA, Fabiano Couto Corrêa da; SALES, Rodrigo de (Org.). **Cenários da Organização do Conhecimento**: linguagens documentárias em cena. Brasília: Thesaurus, 2011.

SOUZA, Renato Rocha; ALVARENGA, Lídia. A web semântica e suas contribuições para a Ciência da Informação. *Ci. Inf.*, Brasília, v. 33, n. 1, p. 132-141, jan./abril 2004.

SOUZA, Sebastião de. CDU: como entender e utilizar a 2ª edição padrão internacional em língua portuguesa. Brasília: Thesaurus, 2010.

WIKIPEDIA. **Região Nordeste do Brasil**. 2015. Disponível em: https://pt.wikipedia.org/wiki/Regi%C3%A3o_Nordeste_do_Brasil. em: 22. Jul. 2021.

WIKIPEDIA. **Região Norte do Brasil**. 2015. Disponível em: https://pt.wikipedia.org/wiki/Regi%C3%A3o_Norte_do_Brasil. Acesso em: 22 jul. 2021.

ZAVAGLIA, Claudia. Ambigüidade gerada pela homonímia: revisitação teórica, linhas limítrofes com a polissemia e proposta de critérios distintivos. **DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada**, v. 19, n. 2, 2003.

ZAVAGLIA, Claudia. **Análise da homonímia no português**: tratamento semântico com vistas a procedimentos computacionais. 360f. 2002. Tese (Doutorado em Letras) - Faculdade de Letras, Universidade Estadual Paulista, São Paulo, 2002.

plo, nenhum estudo rigoroso e compreensivo foi encontrado na literatura da área da Ciência da Informação [...]

Ou seja, urge a necessidade de ação do bibliotecário frente a esse cenário e considera-se o importante papel do profissional da informação com conhecimento em assuntos específicos para possibilitar maior possibilidade de operacionalização de materiais com características peculiares.

Outro ponto que apoia a necessidade de organização e representação da informação musical consiste na grande produtividade de estudos científicos sobre Música nos principais centros de Pós-Graduação no Brasil nos últimos anos, fato percebido devido à vinculação de tais Programas à revistas científicas, Repositórios Institucionais e também na Biblioteca Digital de Teses e Dissertações (BDTD). A finalidade deste vínculo está na maior variedade e visibilidade dos conteúdos musicais, estabelecendo maior divulgação científica, e ainda, com vistas a estabelecer aproximações com outras áreas para o desenvolvimento de soluções para a problemática da recuperação da informação do domínio musical.

As motivações que levaram à sistematização da terminologia desta área do conhecimento, e o conseqüente interesse pela elaboração de um vocabulário são originadas por diferentes ordens: profissional, na Ciência da informação, no sentido de identificar ferramentas de representação da informação musical, bem como da necessidade de instrumentos de controle terminológico nesta área que possibilite a identificação dos itens lexicais recorrentes.

O objetivo dessa pesquisa está em analisar o cenário prático-conceitual da indexação e representação da informação, visando a sistematização, organização de ferramentas de gestão terminológica e recuperação da informação por meio da estruturação de um vocabulário da área de Música.

Concomitante aos propósitos apresentados, a discussão proposta pretende, antes da atribuição de categorias e conceitos/léxicos para construção de um vocabulário, abordar o léxico musical, apontando para a interpretação correta dos termos tratados, situando-os na perspectiva da representação da informação, além de conferir-lhe o respectivo significado e possibilitando relações entre outros conceitos.

REPRESENTAÇÃO DA INFORMAÇÃO E SEUS FUNDAMENTOS

A Ciência da Informação é uma disciplina que investiga as propriedades e o comportamento da informação, as forças que governam seu fluxo e os meios de processamento para otimizar sua acessibilidade e utilização.

Fundamenta-se através da produção, coleta, organização, armazenagem, recuperação, interpretação, transmissão, transformação e utilização da informação. Baseado nesse conceito, Saracevic (1996, p. 47) consegue apreender o enfoque contemporâneo da CI ao redefini-la como:

[...] um campo dedicado a questões científicas e à prática profissional, voltadas para os problemas da efetiva comunicação do conhecimento de registros de conhecimento entre outros seres humanos, no contexto social, institucional ou individual do uso e das necessidades de informação. No tratamento destas questões são consideradas de particular interesse as vantagens das modernas tecnologias informacionais.

Pode-se considerar que a CI foi originada a partir de problemáticas encontradas na sua gênese e evolução no que diz respeito à Documentação e Recuperação da Informação, pontos apontados por alguns teóricos como precedentes desta nova ciência (HARMON, 1971; SARACEVIC, 1992; PINHEIRO, 1997).

Com o crescimento da produção científica, principalmente na área de Ciência e Tecnologia, percebeu-se o fenômeno da “explosão documental”, fato que implicou na preocupação com a organização da informação – onde se destaca trabalhos de pensadores como Paul Otlet e Henri de La Fontaine, percussores na época com a idealização de um controle bibliográfico através de um Sistema de Classificação Decimal Universal (CDU), entre outras metodologias de organização do conhecimento) -, e ainda, na questão do acesso aos documentos com crescimento em grande escala.

Ainda neste período, o advento dos computadores e dos sistemas automatizados nas décadas de 40 e 50, os impactos das novas tecnologias se refletiram na CI assim como em outras áreas do conhecimento. Cientistas perceberam as lacunas em relação ao acesso a informação em crescimento e propuseram formas de solucioná-las, conforme apontava a literatura da época, em especial, um artigo intitulado “*As we may thing*” elaborado por Vanevar Bush, chefe do esforço científico americano na Segunda Guerra Mundial. (BUSH, 1945).

O termo recuperação da informação, cunhado por Mooers (1951) buscava representar “[...] aspectos intelectuais da descrição de informações e suas especificidades para a busca, além de quaisquer sistemas, técnicas ou máquinas empregados para o desempenho da operação.”. Oliveira (2011) corrobora ainda apresentando três questões implícitas no enunciado de Mooers: a) como descrever intelectualmente a informação; b) como especificar intelectualmente a busca; e, c) que sistemas, técnicas ou máquinas devem ser empregados.

A teoria de Mooers visava solucionar substancialmente as necessidades informacionais dos usuários, a partir de estudos sobre o uso da informação e as interações homem-computador. Partindo desse princípio, percebeu-se a necessidade de sistematizar as informações visando a sua recuperação em meios automatizados, por meio dos processos de busca, baseados no uso das linguagens naturais utilizadas pelos usuários. Nesse sentido, os Sistemas de Recuperação da Informação (SRIs) tem o papel de armazenar, organizar, controlar e recuperar os documentos pelos seus conteúdos realmente úteis e o menor índice de documentos não relevantes, visando assim facilitar a comunicação humana. Desse modo, Cesarino (1985, p. 1) apregoa que

Os sistemas de recuperação da informação podem ser definidos como um conjunto de operações consecutivas executadas para localizar dentro da totalidade de informações disponíveis aquelas realmente relevantes. Para isso, executam as funções de seleção, análise, **indexação** e busca das informações.

Desse modo, os Sistemas de Recuperação da Informação (SRI) tratam de procedimentos que organizam, gerenciam e recuperam a informação, em sua maioria ligados a bibliotecas e outras organizações que se focam no gerenciamento da informação visando organizar, recuperar e disseminar a mesma.

Dentre as atividades desempenhadas pelo SRI ilustradas na Figura 1, destaca-se o processo de Indexação por entender que esta é a etapa que está diretamente relacionada com a representação, organização e recuperação da informação nas bases de dados.

A Indexação como atividade técnica segue normas para sua execução, a primeira a ser elaborada foi publicada em 1985 pela *International Standardization for Organization (ISO)*, com a numeração 5.963, intitulada *Documentation – methods for examining documents, determining their subjects, and selecting indexing terms*. A partir de então, a Associação Brasileira de Normas Técnicas (ABNT) publicou a sua tradução e definindo-a como:

[...] a operação que consiste em enumerar os conceitos sobre os quais trata um documento e representá-los por meio de uma linguagem combinatória – lista de descritores livres, lista de autoridades e o thesaurus de descritores – tendo como finalidade a busca documental, que será realizada a partir dos índices ou dos catálogos. (ABNT, 1992, p. 2)

Robredo (2005, p. 165) complementa ainda que “a indexação consiste em indicar o conteúdo temático de uma unidade de informação, mediante a atribuição de um ou mais termos (ou códigos) ao documento, de forma a caracterizá-lo de forma unívoca.”

A partir dessas conceituações, concebe-se Indexação como um processo intelectual de representação do conteúdo de um documento, independente do seu formato, por meio da extração e atribuição de termos visando o aumento do índice de precisão, ou seja, a capacidade de evitar documentos inúteis, e revocação (recall), que significa a capacidade de recuperar documentos úteis no processo de busca manual ou em sistemas automatizados.

A eficiência de um SRI pode ser avaliada a partir da qualidade da Indexação, ou seja, por meio da sua prática com exaustividade, que significa a atribuição do maior número de conceitos para caracterização do conteúdo; e também de acordo com a especificidade do conteúdo, que consiste na exatidão dos conceitos que representam um termo. (LANCASTER, 1968).

O processo de Indexação possui etapas que compreendem desde a análise dos conceitos, bem como a tradução da linguagem natural dos usuários para a linguagem do sistema, a partir das etapas propostas por Lancaster (2004) e reafirmada por outros teóricos da área:

1. Análise do documento e estabelecimento do seu assunto, a partir da leitura das principais partes do documento como: título, resumo, sumário, introdução e conclusões;
2. Identificação dos principais conceitos do documento, e estruturação dos termos que representam o documento;
3. Tradução destes conceitos para os termos de uma linguagem documentária: significando que tais linguagens podem ser alfabéticas, no caso dos cabeçalhos de assuntos e tesouros; ou simbólicas, a exemplo das classificações bibliográficas que utilizam números e letras para representação da informação (DIAS; NAVES, 2007, p. 23).

Convém ainda ressaltar a explosão de informações científicas, que originou a necessidade de uma documentação automatizada, tornando fundamental o uso das contribuições da Linguística para suprir as necessidades da documentação neste contexto. Resgatando os ideais da Semântica sobre significado, as suas contribuições para a indexação são significativas no que diz respeito à pesquisa em processamento automático de texto, em especial, nos estudos de indexação automática onde se buscam soluções baseadas no modelo lingüístico.

Guinchat e Menou (1994) explicam que a indexação automática consiste em “[...] fazer o computador reconhecer palavras que aparecem no título, no resumo do documento, ou no próprio texto. Os termos re-

conhecidos são incorporados em um arquivo de pesquisa e servem para recuperar o documento.". Desse modo, a indexação automática pode ser associada com o uso de *softwares* que permitem identificar a estrutura lexical das frases e o significado dos termos que representam o conteúdo do documento.

Lancaster (2004) apresenta que a Indexação automática se dá por meio de duas tipologias de processos:

1. Indexação por extração automática: Extração de palavras ou expressões do texto para representar seu conteúdo. Pode-se usar um software para extrair os termos a partir dos princípios utilizados por seres humanos (frequência, posição e contexto da palavra no texto);
2. Indexação por atribuição automática: Processo com maior grau de complexidade por se tratar da necessidade de controle terminológico para a representação do conteúdo temático. Desenvolve-se para cada termo atribuído, um "perfil" de palavras ou expressões associativas ao termo e que ocorrem nos documentos.

Percebe-se que as formas de indexação automática apresentadas baseiam-se essencialmente na estatística de palavras de palavras significativas, e ainda, nas contribuições linguísticas por meio da semântica e sintaxe para representação do conteúdo de um documento.

Um aspecto para análise da Indexação consiste nas dificuldades de extrair de forma precisa os assuntos a serem buscados pelos usuários, Mai (2001, p. 606) explica a limitação ao expor que:

Seria quase impossível, naturalmente, para qualquer pessoa ou, neste caso, qualquer indexador, precisar todas as idéias e significados que estivessem associados a qualquer documento, posto que sempre haverá idéias e significados potenciais que diferentes pessoas em diferentes momentos e lugares poderão descobrir nesse documento. Além do que, seria quase impossível prever com exatidão quais das inúmeras idéias e significados que estivessem associados ao documento seriam especificamente úteis para os usuários ou dariam ao documento alguma utilidade duradoura. É de máxima importância reconhecer e aceitar essa indefinição fundamental. O indexador deve compreender, desde o início, que jamais descobrirá todas as idéias e significados que estariam associados ao documento e que, portanto, não é possível descrever todas essas idéias e significados.

Para suprir essa lacuna, alguns fatores possuem influência na exatidão no processo de busca dos termos descritores, ou seja, por meio da adequação de critérios pode-se obter a qualidade da Indexação, apoiada nas considerações de Guinchat e Menou (1994, p. 180), no que diz respei-

to aos pontos: exaustividade; seletividade; especificidade; e uniformidade, porém, Dias e Naves (2007, p. 33) revelam que tais critérios são difíceis de serem operacionalizados.

Tais questões precisam ser observadas visando obter uma indexação orientada para o usuário, conforme adverte Mai (2000, p. 294):

Se focalizar exclusivamente o aspecto da representação, ignorando os usuários futuros, corre-se o risco de representar os documentos de uma forma que não terá qualquer serventia para os usuários. Um indexador que não dê muita atenção nos usuários poderá optar por representar assuntos de documentos que não tenham interesse para eles, ou usar um vocabulário diferente do vocabulário deles, ou representar o assunto em nível que seja muito genérico ou muito específico para eles. No entanto, se o indexador der excessiva atenção aos usuários do sistema, poderá representar os documentos numa forma tal que a representação temática dos documentos somente atenda aos usuários atuais e às necessidades de informação atuais.

Trata-se da observação da linguagem natural utilizada pelo usuário para descrever um determinado assunto; da percepção do indexador para esse aspecto; e a aplicação da linguagem documentária. Compreendem-se as linguagens documentárias como “[...] o conjunto de termos, providos ou não de regras sintáticas, utilizadas para representar conteúdos de documentos técnico-científicos com fins de classificação ou busca retrospectiva de informações [...]” (GARDIN *apud* CINTRA *et al.*, 2002, p. 35), com o intuito de facilitar a comunicação entre a linguagem natural dos usuários e a unidade de informação, bem como representar os conteúdos dos documentos. (TRISTÃO; FACHIN; ALARCON, 2004)

As linguagens documentárias são, primeiramente, instrumentos representativos da informação, atribuídas por meio do tratamento temático da informação, visando estabelecer elos linguísticos com os usuários futuros. Em relação ao percurso histórico, as linguagens documentárias receberam várias denominações, conforme Wanderley (1973) indica, tais como: linguagens descritoras, linguagens de indexação, vocabulários controlados, codificações documentárias, linguagens de informação, lista de assuntos autorizados, linguagem de descrição da informação e linguagem de recuperação da informação.

Com base nessas considerações, principalmente no que diz respeito a estruturação das Linguagens Documentárias, o estudo da linguagem e suas relações torna-se necessários para elencar os termos pertencentes a um vocabulário ou tesaurus, considerados frutos do processo de Indexação visando o controle terminológico e/ou de conceitos.

Kobashi e Francelin (2011, p. 7) mostram que na área da Organização do conhecimento, um conceito pode ter as seguintes características:

- Corresponde a uma unidade de conhecimento;
- É formado por enunciados verdadeiros e verificáveis;
- Não é metafórico, ambíguo ou plurissignificativo;
- Pertence a um domínio de conhecimento.

Tais considerações pautam-se na Lógica e sua formalização, não havendo “[...] preocupação com a semântica em termos dos conceitos e de suas relações; na verdade, todo o foco está centrado em uma dada ‘sintaxe’ que possibilite uma verdadeira ação do pensar. [...]” (CAMPOS, 2004, p. 24).

Atualmente, são desenvolvidos estudos relacionados à análise de conceitos em áreas além da Ciência da Informação como a Linguística, Ciência da Computação e Filosofia, com seus enfoques explicitados no Quadro 2. Cabe ainda ressaltar que nas seguintes áreas, o estudo do conceito também equivale ao estudo de Ontologias.

Quadro 1: Abordagens da análise de conceito por área do conhecimento

Filosofia: representação do ser
Ciência da Computação: recuperação do conceito em sistemas
Linguística: estudo da estrutura do conceito enquanto palavra
Biblioteconomia: estudo do conceito para uso na Indexação

Fonte: a autora.

Em linhas gerais, Medeiros (1986) apresenta que “Conceitos são unidades do conhecimento identificado através de enunciados verdadeiros sobre um item de referência e representados por uma forma verbal (termo ou palavra)”. Lara (2004, p. 237) cita as normas terminológicas e explica que:

A formação de um conceito é considerada fundamental na organização do conhecimento, porque provê os meios necessários ao reconhecimento dos objetos e seu agrupamento em unidades significativas num domínio particular. Esse agrupamento é realizado pela identifi-

ção de características ou propriedades comuns, compartilhadas por tais objetos, ou cuja combinação pode diferir entre culturas, campos ou escolas de pensamento.

Ou seja, a extração do significado, considerado pela autora como objeto, é um elemento principal para a caracterização de um conceito, podendo ser agregado a outro objeto para especificar unidades significativas, consideradas como formação de sentido, e para a sistematização de unidades que podem ser relacionadas a partir de combinações. Estas relações se tornam possíveis de hierarquizações a partir de suas características, que propiciam a determinação do número de funções que exercem como: definição, formação e ordenação classificatória dos conceitos e respectivos índices. (DAHLBERG, 1978).

Enquanto a palavra representa uma unidade linguística significativa ligada à uma classe gramatical e possuem componentes léxicos de uma língua, os conceitos podem ser representados pelos termos, que são designações verbais, e por fim, o termo é considerado a unidade mínima da terminologia, divergindo assim em relação ao significado de palavra, pois o termo por sua vez caracteriza-se como uma palavra contextualizada no discurso (ISO 704, 2000; LARA, 2004). A teórica Cabré (1993, p. 169) aprofunda essa questão ao explicar que:

El conjunto de las palabras especializadas de una determinada disciplina (y también de un ámbito de actividad específica) constituye, como se sabe, la terminología propia de esa especialidad. Los términos, que son las unidades de base de la terminología, designan los conceptos propios de cada disciplina especializada.

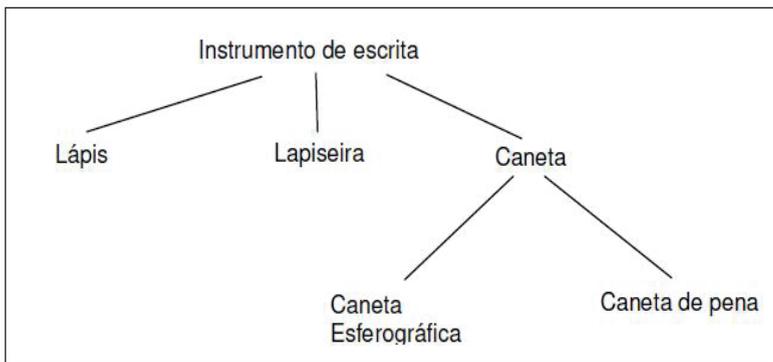
Assim, pode-se inferir que o termo diferencia-se por representar assuntos de áreas específicas, as palavras constituem a unidade que armazena os significados, enquanto os conceitos são definidos pelas características representadas. Dahlberg (1978) assegura que um relacionamento entre conceitos é existente quando dois conceitos possuem alguma característica em comum.

Tais considerações relacionam-se de forma prática com os SRI no processo de Indexação pelo fato de que estes princípios baseiam-se essencialmente na idéia da análise de conceitos sobre determinados assuntos.

Nesse sentido, ao perceber a atualidade de tais considerações, Do-debei (2002) acrescenta que além das relações lógicas, formais, abstratas ou semânticas entre conceitos, devem-se levar em consideração as identidades e diferenças dos objetos, onde se percebe a forte influência dos ideais de Ranganathan, a partir das seguintes relações:

- a) *Relações de equivalência (identidade)*: relacionado à semântica para domínios de dispersão: léxica, com significados parecidos, porém com algumas divergências, e termos com origem linguística diferente e/ou palavras originadas por comunidades específicas pertencentes a uma mesma língua mãe, como as gírias; simbólica, representadas por palavras com grafias diferentes, abreviaturas e nomes complexos, siglas comerciais, traduções; e sintática quando da semelhança entre as características dos conceitos, diferenciados pela forma falada popularmente e de forma científica;
- b) *Relações hierárquicas (implicação)*: por meio da atribuição de uma característica a mais do que outro conceito com as mesmas características, conforme ilustra a Figura 1;

Figura 1: Relações conceituais hierárquicas

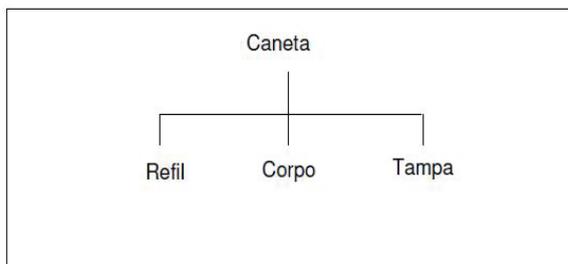


Fonte: ISO 704 (2000, p. 6) tradução

Na figura 2, o termo “instrumento de escrita” apresenta subdivisões hierárquicas, que indicam os conceitos das tipologias dos instrumentos de escrita, e destes, são atribuídas ramificações das suas características, ao impor especificidades com a finalidade de distinguir o máximo possível cada conceito.

- c) *Relações partitivas*: ocorre quando um conceito encontra-se subentendido em outro conceito, permitindo a sua organização em hierarquias lógicas (ANSI/NISO Z39.19, 2003), apresentada na Figura 2:

Figura 2: Relações conceituais partitivas



Fonte: ISO 704 (2000, p.9) tradução

No exemplo citado, a o termo caneta representa o conceito de um todo, já que os termos refil, corpo e tampa representam os conceitos das partes que compõem esse todo.

d) *Relações de oposição (negação)*: quando os conceitos não possuem características em comum, ilustrado a partir das relações de oposição da árvore de Porfírio, como mostra a Figura 3:

Figura 3: Relações conceituais de oposição



Fonte: http://leccionesphilosophica.blogspot.com.br/2010/05/trivium-as-artes-liberais-da-logica_17.html

A Figura 4 apresenta conceitos que não apresentam características comuns. Os conceitos do exemplo possuem características cujas negações encontram-se nas outras características apresentadas.

e) *Relações funcionais (intersecção)*: consideradas também como associativas, consiste na justaposição entre conceitos com elementos em comum, a partir do conhecimento do campo conceitual. Esta tipologia de relação não agrega nenhuma das categorias anteriores, e segundo Currás (1995, p. 155) se estabelecem entre os termos “[...] quando as famílias, ou grupos de termos afins, estudam-se a partir de um plano horizontal, levando-se em consideração diferentes pontos de vista.” A autora expõe ainda que estas relações podem ocorrer nas situações:

- Termos que pertencem à mesma categoria, por exemplo, Termo Geral (TG) Casa – Termo Relacionado (TR) Janela.
- Termos que pertencem à categorias diferentes, elencados do seguinte modo por Araújo (2011):
 - Relação do todo com a parte;
 - Disciplinas ou campos de estudo e os objetos ou fenômenos estudados;
 - Uma operação ou processo e seu agente ou instrumento;
 - Uma ocupação e a pessoa que a exerce;
 - Uma ação e um produto dessa ação;
 - Uma ação e um sujeito paciente;
 - Conceitos relacionados com suas propriedades;
 - Conceitos relacionados com sua origem;
 - Conceitos unidos por dependência de causa;
 - Uma matéria-prima e seu produto;
 - Uma ação e a propriedade associada a ela;
 - Um conceito e seu oposto.

A partir desses pressupostos, é possível visualizar com nitidez as correntes teóricas voltadas para o estudo da linguagem que viabiliza a recuperação da informação e ainda, a construção de linguagens documentárias (LANCASTER, 2004, p. 1). Vale ressaltar ainda que a relação da Ciência da Informação e a Linguística não é uma discussão recente entre os teóricos da CI (BORKO, 1968; MONTGOMERY, 1972; PINTO MOLINA, 1993; MOLLICA, 2012).

A Linguística Aplicada (LA) tem como característica principal a interdisciplinaridade, em uma abordagem recente, como ressalta Praxedes Filho (2012, p. 15):

[...] ciência autônoma que parte de uma visão interdisciplinar da linguagem verbal humana a fim de encontrar soluções para problemas cotidianos relacionados às línguas naturais. Os problemas são apresentados pelos usuários das línguas, sejam eles professores dessas línguas, fonoaudiólogos, tradutores, intérpretes, críticos literários, lexicógrafos,

terminólogos, proponentes de políticas linguísticas etc. As soluções são encontradas pelo linguista aplicado através da pesquisa teórica por ele conduzida.

O autor expõe de forma clara a interseção da LA com outras áreas, porém, cabe aqui ressaltar um dos principais enfoques da LA que tornou este campo ampliado, que consiste no ensino/aprendizagem de línguas estrangeiras, que permite assim os subsídios das pesquisas linguísticas em outros domínios, ou seja, como o próprio nome explana, a aplicação das teorias em questão, visando aperfeiçoar os métodos de ensino da língua.

De nosso interesse, vale ressaltar as especificações fundamentais da microlinguística apresentadas no Gráfico 1:

Gráfico 1: Campos de estudo específicos da Linguística



Fonte: a autora.

Dentre os campos indicados, nos deteremos na Semântica, nos estudos do léxico e Terminologia para fins deste estudo. Em suma, as aproximações entre a Linguística e a CI como foram explicitadas anteriormente, concentram-se nas problemáticas da representação e recuperação da informação, relacionadas diretamente à análise documental, por meio dos subsídios elencados pela análise de conceitos. Nesse sentido, os eixos

que contribuem para solucionar essa questão apoiam-se na Terminologia, nos Estudos do léxico e Lexicografia.

As derivações dos estudos Lexicográficos são os dicionários, glossários e vocabulários históricos, ortográficos e terminológicos. Considerando a linha tênue que interliga a Terminologia, Lexicologia e Lexicografia pelo fato de se constituir de três abordagens distintas sobre o léxico, ressalta-se a relevância de revisitar alguns aspectos e conceitos no que se refere às suas aplicações e produtos. Em síntese, estas considerações são apresentadas no quadro 2:

Quadro 2: Características e objetivos da Terminologia, Lexicologia e Lexicografia

	Características	Objetivos
Terminologia	Estuda o léxico de um domínio, visando a apresentação e transmissão do conhecimento especializado, visando ordenar os termos.	Normalizar
Lexicologia	Formula teorias sobre o léxico, sua estruturação e funcionamento, assumindo assim o caráter científico.	Descrever
Lexicografia	Sistematiza as estruturas linguísticas, sendo considerada uma atividade prática por se ocupar da compilação, classificação, análise e processamento	Estruturar

Fonte: adaptado de (BARBOSA, 1990, p. 154).

A função referencial e normativa são as bases da Terminologia, a normatização com o intuito de nominar e comunicar domínios especializados constitui o seu objeto, que por meio de suas peculiaridades, fornece produtos como os dicionários científicos, glossários, traduções, bancos de dados terminológicos e tesouros.

A partir destes apontamentos, considera-se que a Terminologia, assim como a Lexicografia e Lexicologia distinguem-se pelas suas finalidades, a serem explanadas posteriormente. Ressaltamos aqui o público-alvo de todos estes estudos, que são os especialistas de diferentes áreas do conhecimento, profissionais da informação como bibliotecários, arquivistas, documentalistas, museólogos, linguistas e ainda, os tradutores, intérpretes e redatores técnicos.

A Lexicologia por sua vez possui aproximações estreitas com a Terminologia, que em alguns momentos se torna difícil visualizar suas fronteiras. Andrade (2001, p. 191) explica que “[...] a lexicologia é o estudo científico do léxico, isto é, propõe a estudar o universo de todas as palavras de uma língua.” Com a análise da palavra e do seu conteúdo, além

da sua descrição na Lexicologia, geram-se resultados como a criação de vocábulos e vocabulários, que apoiarão os estudos sobre Lexicografia.

Por fim, a Lexicografia apreende os conhecimentos gerados pela Lexicologia para estruturar a ortografia de um léxico, revelando assim a necessidade da análise prévia dos léxicos para somente a partir desta etapa se realizar a composição de seus instrumentos, sendo assim, a Lexicografia uma atividade posterior a Lexicologia. Turazza (1996, p. 73) explana esta relação ao apregoar que

[...] o lexicógrafo necessita de certos modelos teóricos que expliquem certas características de estruturação de um conjunto lexical, para que possa dar tratamento adequado às unidades lexicais sob seu exame; o lexicólogo, por outro lado, apoiado em dados fornecidos pela lexicografia, pode construir modelos de um universo lexical capazes de permitir a descrição da natureza e das funções deste universo.

Percebe-se que a inter-relação se resume em: a Terminologia é uma sub-área da Lexicologia e Lexicografia que são complementares entre si, estabelecendo assim o que Barbosa (1990) considera como “cooperação recíproca”, sendo que a principal divergência encontra-se na atuação e produtos de cada área.

Além das áreas citadas, convém ainda lembrar que o processamento da linguagem humana em sistemas automatizados, compõe o campo da Linguística Computacional, mostrando assim os elementos da Ciência da Computação que servem para sua composição e contribuições para representação da informação em sistemas por meio de suas relações conceituais.

A Inteligência Artificial (IA), campo da Ciência da Computação, possui abordagens específicas para representação do conhecimento utilizando para isso programas para executar esta função, de forma a se aproximar da inteligência humana. Para tanto, o conhecimento deve estar armazenado de forma que possibilite sua recuperação semanticamente nos bancos de dados, também considerados como bancos de conhecimento, de acordo com alguns autores.

Dentre as principais formas de representação da informação na Ciência da Computação destacam-se a teoria das redes semânticas, com base nos pressupostos de Liu (1994) que representada a informação por algumas estruturas formais e precisas; os *Frames* semânticos que funcionam de forma mais estruturada do que as redes semânticas, e visa com esta forma de representação estender a percepção visual, os diálogos em linguagem natural e outros conceitos complexos (GIRARDI; DRUMOND, 2006), descrevendo seus papéis temáticos e organizando-os em cate-

gorias e trazendo conceituações abrangentes para o termo *frame*; e as Ontologias são formas de representação do conhecimento baseada em ligações semânticas entre conceitos, possibilitando relações com outros conceitos (entendidos também como classes).

Portanto, pode-se pontuar que uso destas estruturas podem ser associadas à representação de léxicos (vocabulários) em linguagem natural, visando o mapeamento das categorias semânticas associadas a um domínio a nível teórico e prático, com a principal função de representar e relacionar terminologias.

Compreendendo a categorização de domínios de especialidade

As contribuições fornecidas pela classificação facetada idealizada por Shiyali Ramamrita Ranganathan em 1933, que preconiza a questão das relações entre conceitos a partir da idéia de que um assunto pode ser categorizado por várias óticas. A ideia mostra-se contemporânea no estudo das relações conceituais a partir dos livros *Prolegomena to Library Classification*, datado de 1937, e ainda, *Elements of Library Classification*, de 1945.

Ranganathan (1967) estrutura cinco ideias (ou categorias) fundamentais, que podem ser compreendidas como “[...] categorias as mais genéricas possíveis e passíveis de se manifestarem de diversas formas, capazes de hospedar todos os objetos da natureza até então conhecidos pelo Homem, e de classificá-los de acordo com sua natureza conceitual, cada um numa e somente numa categoria.”, ou seja, os níveis elencados servem como bases para uma classificação flexível de um determinado domínio, bem como as relações conceituais existentes nas categorias. As Cinco Categorias Fundamentais são apresentadas do seguinte modo:

Há cinco e somente cinco Categorias Fundamentais; são elas: Tempo, Espaço, Energia, Matéria e Personalidade. Estes termos e as idéias denotadas são usadas estritamente no contexto da disciplina de classificação. Não têm nada a ver com seu emprego em Metafísica ou Física. Em nosso contexto, seu significado pode ser visto somente nas declarações sobre as facetas de um assunto – sua separação e sequência. Este conjunto de categorias fundamentais é, em síntese, denotado pelas iniciais **PMEST**. (RANGANATHAN, 1967, p. 398, grifo nosso).

Para o autor, as facetas equivalem a ramificações dentro de um assunto para fins de classificação, sendo atribuídos a cada categoria os seguintes sentidos: Tempo, onde se atribui aspectos como milênio, século, década, ano, etc., características específicas como dia, noite, verão e inverno, e qualidade meteorológica como úmido, seco, tormentoso; Espaço, no que diz respeito aos ambientes geográficos da Terra, dentro e fora

dela, tais como países, estados, etc.; Energia, entendido como a ação de uma espécie sobre a outra, exigindo assim um cuidado maior para sua compreensão, haja vista que esta ação pode ocorrer entre espécies de entidade, inanimada, animada, conceitual, intelectual e intuitiva.

Expõe ainda as facetas: Matéria, sendo que esta categoria subdivide-se em duas espécies, material e propriedade, e consiste basicamente na percepção que as propriedades estão, na maioria das vezes, intrinsecamente ligadas ao material, constituindo-se assim como uma de suas propriedades; e por fim, a Personalidade, que se apresenta indefinível por Ranganathan, cabendo a sua atribuição somente no caso de negação à todas as outras categorias.

Ranganathan mostra ainda que cada categoria de um domínio de conhecimento possui renques e cadeias para organização dos conceitos em estruturas classificatórias hierárquicas. Para o teórico, os renques e cadeias se diferem de acordo com as características de divisão: enquanto os renques formam séries horizontais, a partir de uma única característica, listando assim todos os conceitos relacionados de um aspecto, enquanto nas cadeias cada conceito possui uma característica a mais ou a menos, podendo ser descendente ou ascendente (CAMPOS; GOMES, 2003, p. 161).

Desse modo, a intenção da categorização de Ranganathan incide na organização e sistematização de conceitos, de forma a facilitar a sua recuperação sob várias óticas, para maior refinamento do assunto.

Entende-se a categorização como estruturas de ordenação para fenômenos agrupados e relacionados, com base em suas diferenças, com base na “[...] ocorrência de características comuns em diversos membros, ou seja, o conceito na sua mais ampla extensão.” (ARTÊNCIO, 2007, p. 72).

Isso significa dizer que a categorização nada mais é do que a estruturação de conceitos gerais de um domínio que se relacionam com os conceitos específicos de acordo com seus atributos. O exemplo na Tabela 1 apresenta alguns exemplos de categorizações do domínio musical.

Tabela 1: Exemplos de categorias, subcategorias e termos do domínio de Música

CATEGORIAS	SUBCATEGORIAS	TERMOS/DESCRITORES
Instrumento musical	<ul style="list-style-type: none"> • Instrumentos de sopro • Instrumentos de cordas 	Flauta doce Tuba Violão Guitarra Contra-baixo

Gênero musical	<ul style="list-style-type: none"> • Jazz • Samba • Forró • Rock 	
----------------	--	--

Fonte: a autora

O exemplo mostra que dentro de cada categoria existem várias subordinações de acordo com as características dos conceitos, como por exemplo, na categoria instrumentos musicais, as subcategorias subsequentes podem ser instrumentos de sopro, instrumentos de cordas, etc., de acordo com o exemplo citado. As subcategorias irão abrigar os termos de acordo com o nível de especificidade são vinculados os conceitos pertencentes a cada categoria, fato que reforça a ideia de categoria enquanto conceitos gerais de um domínio.

Em suma, a categorização utiliza o princípio de associação para representar um domínio ou conceito, ou seja, quando a categoria Música é citada para busca, a necessidade específica de informação pode ser relacionada a instrumentos musicais, gêneros, partituras, Etnomusicologia, entre outros aspectos deste domínio.

Por fim, Fujita (2011, p. 64) indica que a estrutura de categorias pode sofrer modificações durante a operação de classificação, e por este motivo é necessário: alterar alguns nomes de categorias em função da especificidade de termos; incluir outras categorias necessárias ao agrupamento de termos; e discriminar as sub-áreas das categorias.

Desse modo, considera-se a categorização como uma forma de organização de SRIs com a finalidade de facilitar a recuperação da informação, sendo consideradas flexíveis e mutáveis de acordo com a abordagem a ser utilizada de um conceito.

PROCESSAMENTO DA LINGUAGEM NATURAL (PLN) E A ORGANIZAÇÃO DO CONHECIMENTO

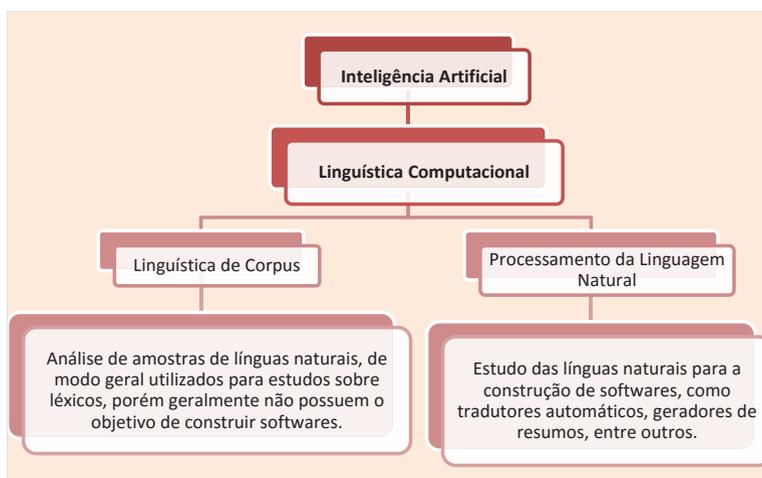
A década de 40 foi significativa para o desenvolvimento dos primeiros computadores, inicialmente utilizados para fins científicos e comerciais de comunicação e armazenamento de dados, haja vista os acontecimentos políticos e militares da época, tendo as suas funcionalidades evoluídas com o passar do tempo. (BRIGGS; BURKE, 2004)

Dentre os vários eixos de estudo da área de Computação, a IA centraliza-se em desenvolver métodos para que os computadores realizem tarefas de modo a simular a inteligência humana na resolução de pro-

blemas. A comunicação e o uso da linguagem em sistemas originaram a necessidade da tradução da linguagem humana para a linguagem de máquinas, utilizada pelos computadores, constituindo assim uma de suas bases de estudo.

Nesta perspectiva, como um ramo da IA surge a Linguística Computacional, que é definida por Vieira e Lima (2001, p. 1) como “[...] a área de conhecimento que explora as relações entre linguística e informática, tornando possível a construção de sistemas com capacidade de reconhecer e produzir informação apresentada em linguagem natural.”. A Linguística Computacional possui duas sub-áreas, que compreendem a Linguística de Corpus e o PLN, com suas ligações hierárquicas e funções descritas no Gráfico 2:

Gráfico 2: Ligações hierárquicas entre as sub-áreas da Inteligência Artificial



Fonte: a autora.

O PLN tem sido estudado pela área da CI na perspectiva teórica, em especial no campo da Indexação e Recuperação da Informação, por entender que os softwares baseados neste modelo propiciam a extração de termos com maior precisão semântica para recuperação da informação em sistemas de busca automatizados. McDonald e Yazdani (1990, p. 176) corroboram com a assertiva de que “[...] a pesquisa em PLN pode proporcionar *insights* bastante úteis sobre processos e representações da linguagem na mente humana, apontando, assim, para a verdadeira IA.”, por utilizar-se dos fundamentos linguísticos sobre os léxicos utilizados pelos usuários nos sistemas de buscas.

Silva (2006, p. 120) sintetiza a evolução dos estudos sobre PLN, bem como os seus efeitos, por ordem cronológica, conforme o Tabela 1:

Tabela 2: Evolução cronológica do Processamento da Linguagem Natural

DÉCADA	ESTUDOS	RESULTADOS
50	Explorações: traduções automáticas	Sistematização computacional das classes de palavras descritas nos manuais de gramática tradicional; Identificação computacional de constituintes oracionais.
60	Formalizações: novas aplicações e criação de formalismos	Primeiros tratamentos computacionais das gramáticas livres de contexto; criação dos primeiros analisadores sintáticos; primeiras formalizações do significado em termos de redes semânticas.
70	Criação do nicho de pesquisa: consolidação do PLN	Implementação de parcelas das primeiras gramáticas e analisadores sintáticos baseados na gramática gerativo-transformacional; busca de formalização de fatores pragmáticos e discursivos.
80	Busca da precisão: sofisticação dos sistemas	Desenvolvimento de teorias linguísticas motivadas pelos estudos do PLN como, por exemplo, a gramática sintagmática generalizada e a gramática léxico-funcional.
90	Busca da precisão e robustez: sistemas baseados em representações do conhecimento no tratamento estatístico de massa de texto	Desenvolvimento de projetos de sistemas de PLN complexos que buscam a integração dos vários tipos de conhecimentos linguísticos e extralinguísticos e das estratégias de inferência envolvidos nos processos de produção, manipulação e interpretação de objetos linguísticos para os quais os sistemas são projetados; ressurgimento da linguística de corpus e do tratamento estatístico de entidades e processos linguísticos.

Fonte: Silva (2006, p. 120)

A tabela 2 mostra que o desenvolvimento dos estudos de PLN ao longo do tempo se deteve inicialmente na tradução da linguagem humana para a linguagem de máquina, e a partir disto, foram agregados os conhecimentos linguísticos, em especial, a implementação da gramática e semântica, entre outros processos em sistemas para a interpretação eficaz dos léxicos utilizados para recuperação da informação.

Em relação aos aspectos conceituais do PLN, também se percebe evoluções em relação às definições de alguns teóricos a partir de suas abordagens que contribuíram para constituição do objeto de estudo dessa área por meio de um encadeamento lógico.

Analisando o PLN, Bobrow *et al.* (1967, p. 161) percebeu que, inicialmente, as preocupações eram direcionadas para o processamento analítico e não estatístico das linguagens naturais, excluindo assim a maioria dos trabalhos em indexação automática, sumarização, análise de conteúdo e de estilo. Desse modo, os modelos de PLN desenvolvidos tinham

seus estudos voltados para a geração de textos e, conseqüentemente, o seu alcance visava contemplar gramáticas ou estudos de línguas específicas em bases de dados, por outro lado, a frequência do uso dos termos não era considerada nesta etapa de estudo.

Mesmo com os debates entre alguns autores acerca da terminologia empregada para esta área – sendo que muitas vezes o PLN confunde-se com a Linguística Computacional, denotando que esta última consiste em uma área principal em que o PLN encontra-se subordinado -, outra ponto debatido nos anos 70 foi a aceitação dos computadores para realização de análises linguísticas, conforme Walker (1973, p. 73) apresenta:

[...] a linguística não aceita o computador ou qualquer computação como um componente metodológico essencial para seu campo, [...] inúmeros linguistas têm questionado não somente a relevância dos resultados das pesquisas em linguística computacional, mas a existência de tais resultados.

Transpostas estas barreiras, a partir do ideal que a Computação, com base nos elementos da Inteligência Artificial e o desenvolvimento de metodologias e recursos computacionais para aprimoramento de técnicas, somente seriam possíveis a partir da interseção com psicologia cognitiva por meio de testes linguísticos para obtenção da manipulação de dados não numéricos (BECKER, 1981).

Somente na década de 80, a estruturação do PLN, com abrangência dos aportes computacionais – em especial, no que diz respeito ao uso de softwares que possibilitaram avanços como: analisadores (*parsers*) de linguagem, representação de significado por computador, entre outros -, foram contemplados conceitualmente, por Warner (1987, p. 79) como “[...] uma área de pesquisa e aplicações que exploram como a linguagem natural, usada como entrada em sistemas de computadores pode ser manipulada e armazenada de forma que preserve certos aspectos do original.”

Chowdury (2003) corrobora ao considerar que PLN é “[...] uma área de pesquisa e de aplicação que explora como os computadores podem ser usados para processar e manipular texto ou discurso em linguagem natural para fazer coisas úteis.” O autor considera ainda que as bases do PLN encontram-se em disciplinas como Ciência da Computação e da Informação, Linguística, Matemática, Engenharia Elétrica e Eletrônica, IA e Robótica, Psicologia, entre outras.

No Brasil, um dos grupos de pesquisa sobre PLN mais expressivos estão o Núcleo Interinstitucional de Linguística Computacional (NILC) vinculado ao Instituto de Ciências Matemáticas e da Computação da Universidade de São Paulo (ICMS/USP) e Universidade Federal de São Carlos

(UFSCAR). Dentre os apontamentos levantados por este grupo, realizou-se o mapeamento das possibilidades de aplicação do PLN em um de seus relatórios, definindo assim o seu aproveitamento nos seguintes processos (SILVA; LIMA, 2007, p. 8): manipulação de bases de dados; sistemas tutores; sistemas de automação de tarefas administrativas; programação automática; sistemas de processamento de textos científicos; sistemas especializados; tradução automática; e sistemas acadêmicos.

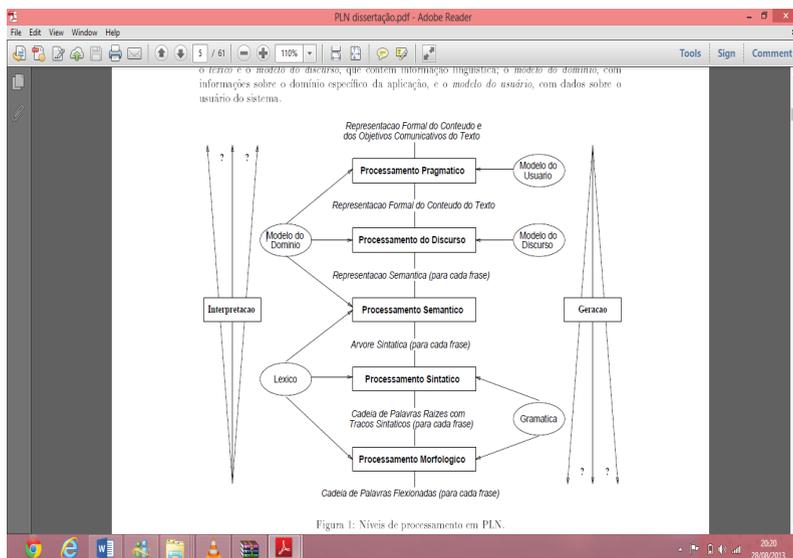
Em suma, a aplicação do PLN refere-se às áreas de: acesso a banco de dados; recuperação da informação; extração da informação; tradução automática e geração de resumos.

Níveis e limitações de PLN

O PLN subdivide-se em níveis de análise e/ou estudo que compreendem: a interpretação, onde são desenvolvidas questões relativas ao estudo da língua de modo que as palavras se tornem compreensíveis pelo computador e, conseqüentemente, o armazenamento para que ocorra a utilização destas palavras em sistemas, tomando como exemplo os tradutores (ou *chatterbots*); e de geração, que ocorre de forma inversa, a partir da inclusão de termos ou expressões, o computador adquire a capacidade de traduzir a compreensão do sistema para a linguagem natural por meio de estruturas semânticas pré-determinadas, no caso dos resumos e palavras-chave.

Tais estruturas fundamentam a arquitetura do PLN, apresentada por Nunes *et al.* (1999) e mostram que o banco de palavras, representado pelo Léxico, é acessado pelos analisadores Léxico, Sintático e Semântico, enquanto a Gramática serve ao analisador semântico para autenticar as palavras ou frases. Nesta perspectiva, o PLN enquanto Sistema baseado no Conhecimento utiliza-se de cinco alicerces: gramática, léxico e o modelo de discurso, ou seja, as informações sobre a língua; modelo de domínio, a ser aplicado; e modelo do usuário que utiliza o sistema (NUNES *et al.*, 1999), ilustrados na Figura 5:

Figura 4: Níveis de processamento em PLN



Fonte: Nunes et al. (1999)

O nível morfológico consiste na definição da estrutura de palavras, bem como a significação e função de cada palavra na frase (adjetivo, substantivo, verbo, etc.); nível sintático, por meio da análise da construção gramatical, suas relações entre unidades linguísticas e sua colocação (sujeito, predicado verbal, etc.); nível semântico, onde as palavras são analisadas pelo seu significado, a partir da análise sintática; nível do discurso, compreensão do significado da palavra a partir do contexto em que ele está inserido; nível pragmático, onde ocorre a compreensão do conteúdo da frase ou texto, a partir da determinação de sua tipologia (pergunta, afirmação) (NUNES *et al.*, 1999).

Constituinte da principal dificuldade do PLN, a ambiguidade, ou seja, a pluralidade de sentidos de uma palavra tem sido uma das motivações para o aprimoramento dos modelos de aplicação do PLN, pois exige uma identificação das unidades gramaticais, a serem aprofundadas nas próximas seções. Dentre as tipologias de ambiguidades encontradas para o PLN, podemos destacar de forma abreviada:

- a) Homonímia lexical, quando uma mesma palavra possui significados que variam na escrita e no pronunciamento. Ex.: "Quem casa quer casa.". A questão encontrada é, em qual dos dois momentos a palavra é verbo ou substantivo?

- b) Ambiguidade sintática, quando a ambiguidade surge a partir da estruturação da palavra na frase. Ex.: “João pediu a Pedro para sair.”, onde a frase gera dois sentidos: 1) João está pedindo para Pedro autorização para sair, ou; 2) João está pedindo para Pedro se retirar.
- c) Ambiguidade de finalidade, de forma resumida significa dúvidas no intuito da palavra no texto. Ex.: “Maria procura um banco.”. A indagação consiste em: o banco que Maria procura é um lugar para sentar ou um banco instituição financeira; neste caso, a solução seria o uso de um analisador de discurso.
- d) Diferentes correferências possíveis, com abrangência em vários níveis linguísticos entre entidades. (RODRIGUES FILHO, 2004).

Nesse sentido, as soluções para esta problemática da ambiguidade estão indicadas no contexto de uso dos termos para assim apreender a sua significação. Com base nisto, revela-se a necessidade da análise linguística em diferentes níveis nas bases de conhecimento, por meio de abordagens que podem ser aliadas ao PLN, tomando como referenciais teóricos metodológicos aspectos morfossintáticos, semânticos e lexicais.

Presume-se que a utilização do PLN enquanto modelo computacional gera a observação de duas características de sua utilização, que diz respeito aos seus aspectos desfavoráveis e favoráveis, sendo que este último fornece uma abrangência maior do que as condições contrárias.

Algumas das limitações do PLN implicam no fato de possibilitar ao usuário a criação de suas próprias consultas, sem ter um padrão a seguir, o que pode ocasionar resultados que não irão satisfazer as expectativas, além de erros na passagem da linguagem natural para uma linguagem de consulta, assim, o usuário pode criar uma resistência ao uso da aplicação ou, até mesmo, não acreditar nas respostas geradas pela mesma (SILVA; LIMA, 2007, p. 2).

Em contrapartida, dentre as vantagens do uso do PLN estão: a eliminação da necessidade de adaptação a formas inusitadas de interação, cuja construção gramatical costuma ser de difícil aprendizado e domínio, a exemplo das linguagens de consulta de bancos de dados (NUNES, 2007, apud NANTES, 2008, p. 26); o usuário não precisa entender o funcionamento de um banco de dados, ele apenas deseja que o resultado da pesquisa seja mostrado de forma simples e objetiva (GARIBA *et al.*, 2005 apud OLIVEIRA; TONIN; PRIETCH, 2010, p. 2); é possível ainda, o entendimento de consulta com erros (termos digitados erroneamente) e incompletas, buscando por palavras próximas e pelo contexto da conversação (SILVA;

LIMA, 2007, p. 2). Para tanto, basta que o usuário tenha um conhecimento básico da área - e ainda, assunto ou domínio -, da especialidade da base de dados.

Softwares de processamento de *corpus*

Na Linguística de *corpus*, o *corpus* de um determinado léxico pode ser analisado por várias óticas e/ou critérios. Candido Junior (2008, p. 32) explica que:

Os softwares para processamento de *cópus* podem ser agrupados em quatro grandes categorias, de acordo com a etapa de construção de *cópus* na qual são usadas: compilação de textos (por exemplo, reconhecedores ópticos de caracteres ou mineradores *Web*), anotação (por exemplo, editores de XML e etiquetadores sintáticos), acesso a *cópus* (por exemplo, concordanceadores e contadores de frequências) e extração de conhecimento (por exemplo, tradutores de máquinas e sumarizadores).

Isso significa dizer que em suma o processamento de *corpus* visa a extração de termos precisos o que nos remete à Terminologia, tendo como resultado principal deste processo, a constituição de glossários, dicionários, vocabulários controlados e tesouros, com auxílio das ferramentas computacionais, que no caso são os softwares extratores.

Quanto a sua constituição, o *corpus* da pesquisa pode ser caracterizado quanto aos seus recursos, que podem ser tanto estruturado, quando se utiliza de dicionários, enciclopédias, glossários, inventários; quanto não estruturado, como livros, teses, artigos, relatórios e entrevistas.

Antes do processamento de *corpus* para extração automática de termos, é necessária a realização de alguns requisitos, para validação e confiabilidade da pesquisa baseada no *corpus*, incluindo se o *corpus* de estudo serve ao propósito inicial da pesquisa. Tais critérios para seleção e compilação de *corpus* são descritos como:

- a) Autenticidade: os textos devem ter sido escrito em linguagem natural, não podendo ser textos “produzidos com o propósito de serem alvo de pesquisa linguística”, e devem ser escritos por falantes nativos (BERBER SARDINHA, 2000);
- b) Representatividade: o *corpus* deve representar características linguísticas da língua sob análise. (SINCLAIR, 1991);
- c) Balanceamento: o *corpus* deve ter um equilíbrio entre os gêneros discursivos (informativos, científico, religioso, entre outros), bem como de tipos de texto (artigo, editorial, entrevista, dissertação, carta, etc.) (ALUÍSIO; ALMEIDA, 2006, p. 159);

- d) Amostragem: o *corpus* deve ter uma amostra proporcional entre os tipos de gêneros e textos selecionados. (BIBER *et al.*, 1998);
- e) Diversidade: o *corpus* deve ter uma diversidade de gêneros e tipos de textos, com a finalidade de abranger uma variação maior de dialetos (BIBER *et. al.*, 1998), e;
- f) Tamanho: o *corpus* deve ter um número total de palavras e categorias suficiente para a sua finalidade (SINCLAIR, 1991).

Ressalta-se ainda que os arquivos que irão compor o *corpus* podem ser originados de documentos impressos, sendo necessário a sua digitalização, e também os que são provenientes da Web, sendo que os arquivos deste último devem ser manipulados antes de seu processamento.

Este processamento de documentos textuais em formato eletrônico passa por algumas atividades, conforme apresenta Aluísio e Almeida (2006, p. 160), como: conversão dos formatos “doc”, “html” e “pdf” para “txt” e limpeza e formatação do *corpus* para o processamento computacional, que trata de remover imagens, gráficos, tabelas, números de páginas e demais anotações, possibilitando a execução de ferramentas como contadores de frequência, concordanciador e ferramenta de extração automática de termos. Após a execução destas etapas, o *corpus* deve ser nomeado, gerando cabeçalhos que possibilitem a sua recuperação futura pelos softwares de processamento de *corpus*.

De acordo com Bagot (1999), a metodologia para utilização de termos ou sistemas extratores de candidatos a termos classificam-se em três categorias: sistemas baseados em conhecimento estatístico; sistemas baseados em conhecimento linguístico e; sistemas híbridos.

Entende-se por softwares estatísticos, aqueles que empregam os dados de frequências de ocorrências de elementos lexicais, para assim extrair os termos que representam o documento em questão. Citamos na Tabela 3 os exemplos deste tipo, de natureza gratuita:

Tabela 3: Exemplos de softwares extratores estatísticos

SOFTWARES	CARACTERÍSTICAS/FUNCIONALIDADES
Pacote NSP (N-gram Statistics Pack-age)	Realiza a identificação e extração de termos (n-gramas), ou seja, de sequências de caracteres de comprimento, que podem ser unigramas, bigramas, trigramas e tetragramas. Com abordagem puramente estatística, para a sua execução é necessário do software Perl instalado, e por não possuir interface gráfica, funciona via linha de código.

Corpógrafo	Criado pelo Centro de Linguística da Universidade do Porto, para criação de corpora específicos próprios, com capacidade de extração terminológica e de preparação de definições. Dentre suas funções, pode-se destacar: pesquisar nos corpora utilizando expressões regulares, criar listas de N-gramas, obter listas de candidatos a termos, criar novas classificações para domínios específicos, relações semânticas e conectores de discurso, visualizar as ocorrências de termos nos corpora, visualizar redes lexicais constantes numa base de dados. (LINGUATECA, 2014).
ZExtractor	Criado por José Lopes Moreira Filho em 2009, com a finalidade de atender à demandas de extração automática a candidatos de termos, utilizando o recurso das palavras-chave. O ZExtractor possibilita o ajuste parâmetros estatísticos; extração precisa de n-gramas; interface gráfica na qual é possível o usuário definir um número mínimo de ocorrências para que uma palavra seja candidata a termo; e o estabelecimento dos itens que devem ser excluídos, ou seja, os stoplists.

Fonte: adaptado de (TEIXEIRA, 2010)

Os softwares indicados da Tabela 3 apresentam abordagem estatística com base na extração de n-gram, ou seja, de acordo com a extensão de unidades lexicais. Pazienza *et. al.* (2005) explica que existem mais medidas estatísticas que podem denotar relações entre as que podem ser utilizadas na extração de termos, que representam a dimensão linguística dos termos extraídos. Outro método que visa aprimorar a extração de candidatos a termos tem como base a Linguística e suas propriedades de processamento.

Os softwares linguísticos possuem a funcionalidade de identificar informações linguísticas baseando-se em conjuntos de anotações linguísticas como análise morfológica, morfossintática, sintática, semântica e pragmática.

Para Dalmolin (2010, p. 38) “Na abordagem linguística são utilizadas técnicas de análise de discurso, processamento de linguagem natural e reconhecimento de padrões previamente definidos por projetistas que geralmente são especialistas no domínio.” Alguns exemplos de softwares linguísticos são apresentados na Tabela 4:

Tabela 4: Exemplos de softwares extratores linguísticos

SOFTWARES	CARACTERÍSTICAS/FUNCIONALIDADES
WebCorp	Desenvolvido pela Research and Development Unit for English Studies (RDUES) na School of English da University of Central England, em 2000. Consiste em um conjunto de ferramentas que permitem acesso a Web como um recurso linguístico, ou seja, realiza a extração de vários aspectos sobre línguas como se a Web fosse um corpus. Possui como público alvo linguistas, lexicógrafos, editores, jornalistas, pesquisadores, professores de língua que estudam o uso da língua, seus neologismos, entre outros.

Unitex	Implementação livre do programa Intex, desenvolvidos pelo laboratório francês Laboratoire d'Automatique Documentaire et Linguistique (LADL). O software Unitex consiste em um conjunto de programas para processamento de corpus linguísticos com interface Java, que permite que a ferramenta não atrapalhe o desempenho de outras plataformas durante o processamento de corpus. Os principais recursos linguísticos do Unitex são: dicionários (para serem utilizados pela máquina e não para humanos), e tabelas do léxico-gramática, que consistem em matrizes binárias que mostram as propriedades de algumas palavras.
GATE (General Architecture for Text Engineering)	Desenvolvido pelo Grupo de Processamento de Linguagens Naturais do Instituto de Ciência da Computação da Universidade de Sheffield, na Inglaterra. O software foi desenvolvido em código aberto (livre), baseado em Java, com a finalidade de solucionar as problemáticas que envolvem a análise e processamento de texto, como extração de informações por meio de construção de taxonomias via menus, etiquetagens morfossintáticas e anotações semânticas e tratamento de co-referência ou anáforas. Os principais recursos disponibilizados pelo GATE são: Visual Resources (VR), representam a visualização e edição de componentes; Language Resources (LR), que representa entidades tais como dicionários, corpora ou ontologias, ou seja, entidades léxicas de um domínio; Processing Resources (PR) representa entidades que são primeiramente algorítmicas de PLN, como parsers, geradores ou modeladores do sistema ANNIE (Nearly-New Information Extraction).
LácioWeb	É um projeto financiado pelo Conselho Nacional de Pesquisa (CNPq) em parceria com entre o NILC e a Faculdade de Filosofia, Letras e Ciências Humanas (FFLCH/USP), possui ferramentas como contador de frequência padrão; contador de frequência por palavra; concordanciador para corpus sem anotação; etiquetadores morfossintático; editor de cabeçalhos, entre outros. Em suma, o intuito do Lácio-Web é de representar corporas devidamente compilados, catalogados e codificados por um padrão que permita o intercâmbio, navegação e análise.
SYNTEX	Ferramenta chamada de analisador de corpus para extração de termos em corpus de língua francesa, que utiliza textos anotados por um parser para construção de um léxico específico do domínio e paralelamente realiza a análise sintática. A extração de termos é feita por meio de sintagmas nominais, levando em consideração as categorias morfossintáticas e as principais relações sintáticas como sujeito, objeto direto e complemento proposicional.

Fonte: adaptado de (ALUÍSIO; ALMEIDA, 2006; LÁCIO-WEB, 2014; LOPES; VIEIRA, 2010)

A Tabela 4 indica as funcionalidades da lingüística computacional aplicada em softwares, por meio do uso de ferramentas fundamentadas em análises semânticas, morfossintáticas, anotações e/ou etiquetagem, entre outros, para refinamento dos termos processados.

Pazienza *et. al.* (2005) indica que a abordagem linguística na extração de termos deve: analisar o corpus de um domínio e identificar a *Parts of Speech* (PoS) – que são as categorias sintáticas ou gramaticais -; identificar e extrair termos a candidatos conforme as regras linguísticas criadas; preservar os significados de acordo com o termo original e; implementar filtros linguísticos para refinar a terminologia, fatos estes que requerem

um tempo maior para seu desenvolvimento, para que o refinamento seja mais preciso.

E por fim, a integração das funções estatísticas e linguísticas para extração de termos são características dos softwares híbridos, que combinam os modelos de frequência de ocorrências, com a base linguística, e por este motivo, tende a melhorar os resultados devido o equilíbrio entre cada abordagem.

Pazienza *et. al.* (2005) apresenta que esse tipo de abordagem tende a alcançar resultados mais corretos se comparado às abordagens puramente estatística ou puramente linguísticas. Deste segmento, destacamos de forma breve, um software híbrido que é voltado para construção de ontologias:

- **OntoGen:** é uma ferramenta para extração automática de candidatos de termos, que identifica os documentos que correspondem ao tema e a seleção pode ser refinada pelo computador do usuário, e realiza a determinação de hierarquia de conceitos. (LOPES; VIEIRA, 2010, p. 193).

A partir destas considerações iniciais, vale ressaltar que o PLN não é um modelo de recuperação da informação, e sim um método de interação que pode ser efetivado em sistemas de informação (ou bancos de dados específicos) visando interpretar de forma mais precisa possível a linguagem dos usuários, focando o texto, uma vez que as expressões utilizadas para busca da informação são constituintes dos objetos linguísticos.

O DOMÍNIO MUSICAL COMO MEIO DE EXPRESSÃO

A Musicologia, isto é, ciência que estuda a Música em suas teorias, surgiu na segunda metade do século XIX, próximo ao acontecimento da I Guerra Mundial, conforme apresenta Mello (1997), com influências matemáticas e filosóficas da época. Atualmente, torna-se indispensável tratar de Musicologia quando se objetiva compreender qualquer manifestação musical.

A partir de então, o musicólogo austríaco Guido Adler (1985 *apud* DUDEQUE, 2004) dividiu a Musicologia como ciência em dois ramos: histórica e sistemática, ao apregoar suas divergências do seguinte modo: a Musicologia Histórica busca apreender os fenômenos ligados à evolução da música em relação às suas produções documentais, tais como as composições e formas de acordo com cada época, enquanto a Musicologia

Sistemática aprofunda questões relativas às componentes da música em si e dos sons, tais como harmonia, ritmo, melodia.

Como marco teórico na Idade Média, apoiado nas sistematizações das ciências por meio dos modelos pedagógicos do *trivium* que Granja (2006, p. 41) explica que era composto pelas disciplinas da linguagem: gramática, retórica e dialética e *quadrivium* que acrescentava a aritmética, geometria, astronomia e música, e juntos formavam as sete artes liberais da Grécia, sendo referências curriculares do ocidente por mais de 1.000 anos.

Na perspectiva filosófica educacional as influências das Artes, em especial a Música, tiveram seus reflexos registrados na história. Na Grécia, a Música era diretamente relacionada à Filosofia e à Educação, acreditava-se que os seus efeitos agiam diretamente sobre a mente, corpo e alma, e por estes motivos, eram restritas somente aos cidadãos livres.

Granja (2006, p. 38) mostra a presença da Música nos conceitos do filósofo Platão na doutrina do *éthos* musical, ao representar que tais conceitos agregavam valores éticos e estéticos à alma, e explica: “Os gregos chamavam de “*éthos* musical” o caráter particular associado a um determinado modo musical. Assim, um modo poderia exprimir um *éthos* do homem valente ou do homem sereno, enquanto outros estariam associados aos maus hábitos, à preguiça, à paixão.” (GRANJA, 2006, p. 38).

Trata-se de um elemento antropológico-cultural, que embora se origine e se desenvolva na esfera dos sentimentos, das emoções, do gosto pessoal, da sensibilidade e da subjetividade, tem também uma objetividade, cujas fontes se encontram no ambiente natural, histórico e social do povo. Para Queiroz (2000, p. 17)

O conteúdo musical trata da mensagem, de caráter emocional, presente na música. Ou dizendo de outro modo, o conteúdo musical é aquilo que a música transmite, o estado que a música porta. E, por mais constatação que possa haver quanto à definição do que ela porta, deve ser claro que algo ela porta. [...] quando verdadeiramente artístico, dá testemunho da verdade e da harmonia possível a vida – em uma forma compreensível à sensibilidade emocional.

Assim, pode-se afirmar que a Música reflete o que somos, o nosso modo de ser, de pensar as coisas, de relacionar-se com as pessoas e com o universo, ou seja, “[...] a música é uma arte eminentemente social, portanto, vinculada a sua época e ao seu lugar, suscetível às variações da sociedade incluindo evolução tecnológica.” (FERREIRA, 2001, p. 92).

No campo da Antropologia, a Música é estudada como Etnomusicologia, que visa a inter-relação entre os aspectos históricos e culturais

entre estas duas áreas, a exemplo disso, o antropólogo americano Alan P. Merriam (1964, p. 27) estabeleceu a teoria da Etnomusicologia e apregoa:

A música é um fenômeno exclusivamente humano que só existe em termos de interação social; ela é feita por pessoas para outras pessoas, e isso é um comportamento aprendido. A música não existe por, de ou para si mesma. É preciso haver sempre os seres humanos a fazer algo para produzi-la. Em suma, a música não pode ser definida como um fenômeno do som sozinho, pois envolve o comportamento de indivíduos e grupos de indivíduos, e suas organizações particulares, exigências, e a concordância social das pessoas que a compõem.

O autor expõe que a Música constitui-se como um produto da cultura de um povo. Hummes ressalta dez funções da Música na Etnomusicologia, apresentada por Merriam, a saber: 1) expressão emocional; 2) prazer estético; 3) divertimento; 4) comunicação; 5) representação; 6) reação física; 7) impor conformidade às normas sociais; 8) validação das funções sociais e dos ritos religiosos; 9) contribuição para a continuidade e estabilidade da cultura e; 10) contribuição para a integração da sociedade. (MERRIAM, 1964, apud HUMMES, 2004)

Estes elementos são resultados de processos e interpretações sociais, representados em diferentes formas. A Música é compreendida em seu contexto cultural e sua análise é feita de forma cognitiva. Nesse sentido, a preservação da Etnomusicologia mostra-se como elemento constituinte da memória, tradição e expressões artísticas de um lugar.

Por meio destas considerações, entende-se que a Música além de uma expressão artística, é entendida também como uma forma de linguagem, capaz de exprimir as realidades cotidianas, fato este que justifica as expressões musicais próprias de cada civilização ou povo, caracterizando-se como um instrumento de identidade e transformação social.

DESCRIÇÃO METODOLÓGICA

A compreensão e interpretação dos fenômenos a partir do contexto em que estão inseridos são fatores integrantes na produção do conhecimento. Para isso, é necessário o emprego de métodos para realização da pesquisa científica. Demo (1996, p. 34) aponta que a pesquisa é uma atividade cotidiana, considerando-a como uma atitude, um “[...] questionamento sistemático, crítico e criativo, mais a intervenção competente na realidade, ou diálogo crítico permanente com a realidade em seu sentido teórico e prático.”

O estudo trata-se de pesquisa aplicada (gerando aplicações práticas, dirigidas para problemas específicos), de natureza teórico-exploratória. Teórica, uma vez que investigamos em livros, artigos e afins a problemá-

tica da construção de vocabulários, genéricos e especializados, com base no tratamento computacional de dados linguísticos, seja por meio do PLN e suas ramificações, como: recuperação da informação, motores de busca, etiquetadores, desambiguadores, entre outros.

E ainda, de natureza exploratória ao propor de um vocabulário de um domínio de conceitos, com base no agrupamento de itens lexicais especializados, no caso, do campo musical. O percurso metodológico adotado para a realização da pesquisa norteia-se pelo procedimento de pesquisa bibliográfica e documental sobre as temáticas das áreas de Ciência da Informação, Linguística, Computação e Música, e por este motivo, assume o caráter de interdisciplinar, ao aplicar estes campos no modelo de pesquisa em Processamento Automático de Línguas Naturais (PLN).

O ambiente e a constituição do corpus da pesquisa

A estruturação de um *corpus*, ou seja, a definição de “[...] uma coleção de documentos coletados dentro de determinados padrões ou exigências, para a realização de estudos linguísticos ou computacionais de aprendizagem de máquina.” (SILVEIRA, 2008, p. 29), ou seja, uma amostra do léxico, que irá compor o objeto da pesquisa em formato eletrônico, sendo especializado por se tratar de um domínio do conhecimento.

O *corpus* de Música envolveu etapas definidas, como levantamento de teses e dissertações referente ao domínio de Música, produzidas nos principais Programas de Pós Graduação em Música dos Centros/Instituições no Brasil, além de artigos científicos de revistas deste campo:

- Universidade de Campinas (UNICAMP);
- Universidade Federal do Rio Grande do Sul (UFRGS);
- Universidade Federal do Paraná (UFPR);
- Universidade Federal do Rio Grande do Norte (UFRN);
- Universidade Federal da Bahia (UFBA);
- Escola de Música (UFMG);
- Biblioteca Digital de Teses e Dissertações;
- Revista eletrônica de Musicologia;
- PERCEPTA – Revista de cognição musical;
- Música em perspectiva;
- Per Musi – Revista acadêmica de Música;
- Música e cultura – Revista da Associação Brasileira de Etnomusicologia;

- Música em contexto;
- Revista Opus.

A definição da amostra compreende o período de 2003 à 2013, com *corpus* não estruturado, dentre os quais foram selecionados o quantitativo de 10 dissertações, 10 teses e em média 30 artigos - haja vista que a maioria das revistas possuem 2 publicações anuais -, produzidos por ano em todos os programas e/ou revistas citados. Vale ressaltar que tais documentos foram obtidos em formato eletrônico e em pdf, por meio de pesquisas na Web nos repositórios digitais dos Programas/Revistas citados.

A escolha dos *corpora* dos Programas de Pós Graduação e Revistas citadas pautaram-se no fato de se tratar dos principais centros e veículos de comunicação científica do país voltados para a Música que disponibilizam suas produções em formato eletrônico em seus repositórios institucionais, se tornando assim fontes de literatura especializada.

A seleção justifica-se ainda por tais Programas/Revistas se constituírem como fonte de coleta tanto dos conceitos quanto dos itens lexicais recorrentes no domínio de Música, contribuindo assim para solucionar a problemática da inexistência de instrumentos de controle terminológico para este domínio.

Os aspectos práticos da pesquisa foram executados diretamente pela plataforma E-terms, que oferece em um dos seus módulos a extração automática de unidades ou conjuntos lexicais que podem constituir uma unidade terminológica, ou seja, um termo, baseado em operações estatísticas proporcionadas pelo pacote NSP statistic Pack-age, com funções integradas no E-terms.

Desenvolvido em parceria com a Embrapa Informática Agropecuária, Universidade de São Paulo (USP Campus de São Carlos) e a Universidade de São Carlos (UFSCar), o e-Terms constitui-se como um ambiente colaborativo web de gestão terminológica. Tem como objetivo integrar e viabilizar criação de produtos terminológicos por meio da (semi) automatização das etapas do trabalho terminológico e também reúne todas as etapas do trabalho de criação dos produtos terminológicos em uma única arquitetura computacional.

Dentre suas funcionalidades estão: compilação automática e semi-automática de *corpus*; ferramentas de análise da qualidade de *corpus* como contadores de frequências de palavras, contadores de frequência de uma única expressão; identificação e recuperação de termos; etiquetadores; editor de ontologia integrado; exportação de produtos termino-

lógicos; ambiente web de acesso livre e gratuito. Para tanto, o e-Termos divide as atividades colaborativas a partir de 6 etapas definidas, a saber:

- Compilação automática de corpus;
- Suporte e análise da qualidade dos corpora;
- Extração automática de termos;
- Edição do mapa conceitual e categorização de termos;
- Gerenciamento da base de dados terminológicos;
- Intercâmbio e difusão de termos (E-TERMOS, 2014).

Vale ressaltar que este ambiente computacional possibilita a utilização integrada do Pacote NSP para a extração automática de termos no módulo/etapa 3 do e-Termos, que será utilizado para fins deste estudo.

Extração automática de termos

A sistematização do vocabulário de Música na pesquisa abrange o uso do ambiente colaborativo de gestão terminológica, e-Termos, em especial, no que diz respeito a utilização de uma funcionalidade de extração de termos deste ambiente, viabilizada pelo uso do *software* estatístico, o Pacote *N-gram Statistic Package* (NSP) em sua interface.

Desse modo, a extração automática a candidatos de termos tem com base o Processamento da Linguagem Natural, visando maior extração do conhecimento semântico dos textos processados. As etapas de fundamentação metodológicas são:

- a) Busca e seleção de fontes não estruturadas, no caso, teses e dissertações disponíveis em formato eletrônico e em pdf nas bases de dados definidas;
- b) Compilação do *corpus*: esta etapa envolve o armazenamento do *corpus*;
- c) Manipulação dos arquivos do *corpus*;
- d) Inclusão dos textos do *corpus* no e-Termos;
- e) Levantamento e análise da lista de unigramas, bigramas, trigramas, que correspondem a termos compostos por uma, duas ou três unidades, respectivamente, realizado pelo software Pacote NSP, integrado ao e-Termos;
- f) Limpeza das listas geradas, com eliminação de unidades que não correspondem a termos;

- g) Validação: escolha de julgadores, pertencentes a área abordada pelo vocabulário para que “[...] o repertório da linguagem documental represente de maneira mais real e mais precisa possível a forma como determinada comunidade busca a informação desejada.” (MORAES; CRISTIANINI, 2010); e preparação do material a ser conduzido para os julgadores para definição dos critérios para escolha dos termos definitivos;
- h) Identificação das categorias;
- i) Organização e apresentação do vocabulário de Música.

PROCESSAMENTO DA LINGUAGEM NATURAL DO DOMÍNIO MUSICAL

O objetivo desta seção é apresentar a metodologia de desenvolvimento do vocabulário musical, que utiliza o pacote NSP para realização da extração estatística de candidatos de termos. Devido o seu funcionamento ser via linha de comandos, o E-terms criou um ambiente com interface para o usuário para auxiliar na construção de vocabulários baseada em PLN.

A abordagem utilizada é considerada semi-automática, devido à intervenção humana, que abarca quatro etapas principais: 1) construção de um vocabulário do Domínio Musical por meio da Indexação manual; 2) Compilação e processamento automático do corpus no ambiente E-terms; 3) Extração automática de candidatos a termos; e 4) Cálculo da frequência dos termos candidatos ao vocabulário, por meio de tarefas manuais e automáticas.

Sistematização preliminar do Vocabulário Musical

A leitura técnica preliminar do *corpus*, em especial de informações contidas no título, palavras-chave, resumo, título das seções, introdução e conclusão, torna possível a extração manual de candidatos a termos, com base na técnica da Indexação, haja vista que tais informações embasarão o vocabulário do domínio musical.

A motivação por este tipo de procedimento está, principalmente, no registro destas informações, entendidas aqui como categorias, ou seja, possíveis termos descritores, e itens lexicais, que posteriormente serão comparadas aos termos que serão extraídos automaticamente, visando analisar as aproximações dos resultados obtidos.

A estruturação do *corpus* foi feita a partir de etapas pré-definidas, como a busca em bases de dados de fontes não estruturadas específicas, para compilação (armazenamento) dos arquivos obtidos. Após esta

etapa realizou-se a manipulação do *corpus*, que consiste na “limpeza” do texto nos formatos “doc”, “HTML” e “pdf”, transformando-os em formato “txt”, também conhecido como bloco de notas, removendo todos os números, gráficos e/ou imagens do arquivo, tornando-o puramente textual com anotação e formatação, tornando o *corpus* mais manipulável para processamento computacional.

Após a indexação manual, fase que prepara a lista de termos em que a extração se baseia, e retoma-se o ideal de categorização a partir dos 699 textos, sendo estes, 358 dissertações, 100 teses e 241 artigos científicos indexados manualmente, que geraram o quantitativo de 29 categorias conceituais que abrigam as unidades lexicais que representam o Domínio Musical.

As categorizações obtidas e os itens lexicais, sistematizados em ordem alfabética estão dispostos no Apêndice A, conforme a ocorrência dos termos nos textos e após a validação por especialistas do domínio de Música. A partir desta estruturação foi produzido também, o Índice de termos do domínio musical, apresentado no Apêndice B, para facilitar a identificação dos conceitos obtidos em ordem alfabética.

A elaboração de um *corpus* computadorizado obedece a critérios, preconizados por Sinclair (1991, *apud* BERBER SARDINHA, 2000) como a representatividade, ou seja, a extensão do *corpus* que segundo o autor, deve ser o maior possível para obter representatividade. Para tanto, a título de informação, Berber Sardinha (2003) apresenta a classificação geral de tamanho de *corpus*, indicada na Tabela 5:

Tabela 5: Classificação de tamanho de corpus

TAMANHO EM PALAVRAS		CLASSIFICAÇÃO
Menos de 80 mil	Pequeno	
80 a 250 mil	Pequeno-médio	
250 mil a 1 milhão	Médio	
1 milhão a 10 milhões	Médio-grande	
10 milhões ou mais	Grande	

Fonte: (BERBER SARDINHA, 2003).

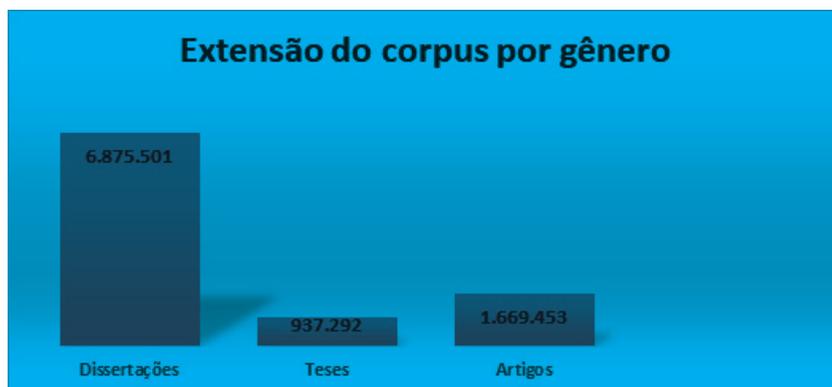
Nessa perspectiva, ao final do processo compilatório, obteve-se um *corpus* médio-grande constituído por 9.482.246 palavras, extraídos de 424 textos selecionados para o processamento de texto, extraídos de 14 fontes diferentes, entre produções técnico-científica de Programas de Pós Graduação e revistas científicas. Vale ressaltar que o quantitativo inicial de textos que participaram da indexação manual foi de 699 textos, com os

cortes de 275 textos, tendo assim, como produto final 424 textos participantes.

Em termos quantitativos para fins de processamento automático pela plataforma E-terms, o corte citado justifica-se pela necessidade da exclusão dos textos que tiveram problemáticas tanto no momento da conversão do formato PDF para txt, para processamento do texto. Os documentos digitalizados também perpassaram pelo mesmo entrave, fato que, no caso do gênero Teses, justifica a redução de textos participantes no processamento de textos.

Na etapa 2 do E-terms, a opção contador de palavras permite ao usuário medir o quantitativo de palavras do corpus analisado. Dividimos a extensão de acordo com o gênero técnico-científico adotado que compõe o *corpus* do domínio musical, conforme apresenta o Gráfico 3:

Gráfico 3: Extensão do corpus por gênero



Fonte: dados da pesquisa.

Desse modo, os dados revelam que o tamanho do corpus obtido sobre o domínio musical enquadra-se na classificação grande, por apresentar um total de 9.482.246 palavras. Na operacionalização do E-terms, a ferramenta “Contador de frequência” da segunda etapa, apresenta ainda dados como total de palavras diferentes e índice de riqueza vocabular, e os resultados são:

- Dissertações, com um total de palavras diferentes de 197.339, e índice de riqueza vocabular de 0.03;
- Teses, com 46.311 palavras diferentes e 0.049 de índice de riqueza vocabular, e;
- Revistas, com um total de 77.077 palavras diferentes e índice de riqueza vocabular de 0.046.

Estas considerações indicam, além do quantitativo de palavras de cada corpus, o nível de representatividade e riqueza vocabular de termos, válidos para construção e análise do vocabulário controlado do domínio musical.

Vocabulário de Música: estruturação e análise

Nesta seção, apresentam-se os resultados obtidos com a extração automática de termos do corpus e sua representação como unidades lexicais de um vocabulário sobre o Domínio Musical.

A pesquisa utilizou as ferramentas da plataforma colaborativa E-termos, apresentada em seções anteriores, que fornece acesso livre e gratuito. Para ter acesso é necessário a realização de um cadastro, e em seguida o pesquisador precisa propor um projeto para utilizar as ferramentas disponíveis. Após a aprovação da proposta, tem-se a opção de abrir o projeto para operacionalizar as etapas dispostas pelo ambiente, conforme a Figura 7:

Figura 6: Página de identificação do projeto do usuário do E-termos

The screenshot shows the e-Termos web interface. At the top, there is a teal header with the text "e-Termos Ambiente Colaborativo Web de Gestão Terminológica." Below the header is a navigation menu with links for "Cadastro", "Projeto", "Convite", "Sobre", and "Sair". The main content area displays the user's name "Olá Juliana Rabelo do Carmo" and a message: "Estes são os Projetos e respectivos Perfis que você atualmente participa no e-Termos. Para acessar um determinado Projeto, verifique o Perfil desejado e clique em **Abrir Projeto**." To the right of the message is an icon of a folder with a star. Below the message is a table with the following data:

N.	Nome do Projeto	Perfil	Ação
1	TÉCNICAS E TECNOLOGIAS DE EXTRAÇÃO AUTOMÁTICA DE TERMOS NO DOMÍNIO MUSICAL	Gerente de Projeto	<input type="button" value="Abrir Projeto"/>

Below the table, there is a footer with the text "EMBRAPA/CNPq - NILC/USP - GETerm/UFSCar - Condições de Uso" and "Projeto e-Termos - Todos Direitos Reservados - 2009". To the right of the table, there is a button labeled "Propor Novo Projeto".

Fonte: <https://www.etermos.cnptia.embrapa.br/main1.php>

Após clicar na ação de "Abrir Projeto", o usuário será direcionado para a página inicial, ilustrada pela Figura 8, de execução das etapas - já mencionadas anteriormente -, do E-termos, que para fins desta pesquisa utilizaremos as etapas 2 que consiste no suporte e análise da qualidade dos corpus, e 3 que trata da extração automática de termos.

Figura 7: Tela principal do E-termos

e-Termos Ambiente Colaborativo Web de Gestão Terminológica

Principal | Etapa 1 | Etapa 2 | Etapa 3 | Etapa 4 | Etapa 5

Projeto **TÉCNICAS E TECNOLOGIAS DE EXTRAÇÃO AUTOMÁTICA DE TERMOS NO DOMÍNIO MUSICAL** **Principal**

Perfil **Gerente de Projeto**

Projeto	Etapa	Equipe	Usuários	Mail	Recado
Status das Etapas deste Projeto					
Etapa	Status Atual				
Etapa 1	Liberada				
Etapa 2	Liberada				
Etapa 3	Liberada				
Etapa 4	Bloqueada				
Etapa 5	Bloqueada				
Etapa 6	Bloqueada				

[Clique aqui](#) para mais informações sobre as Etapas de trabalho do e-Termos.

Fonte: <https://www.etermos.cnptia.embrapa.br/main2.php>

Utilizamos a etapa 2 para fazer o upload dos textos que irão compor o *corpus* da pesquisa, que pode ser carregado de forma individual ou em pasta com extensão zipada, ou seja, os documentos ficam agrupados em um único arquivo. É necessário ainda estabelecer o gênero e o tipo textual dos textos que irão ser compilados para o *corpus*, como mostra a Figura 9.

Figura 8: Segunda etapa do E-termos – Suporte e análise da qualidade dos córpus

e-Termos Ambiente Colaborativo Web de Gestão Terminológica

Principal | Etapa 1 | Etapa 2 | Etapa 3 | Etapa 4 | Etapa 5 | Etapa 6

Projeto **TÉCNICAS E TECNOLOGIAS DE EXTRAÇÃO AUTOMÁTICA DE TERMOS NO DOMÍNIO MUSICAL** **Segunda Etapa**

Perfil **Gerente de Projeto**

Projeto	Recado	Mail	Textos	Córpus	Ferramentas	Fórum
Preencha os campos abaixo e clique em Enviar .						
Upload de Textos						
Arquivo:	<input type="text" value="Escolher arquivo"/>	<input type="text" value="CORPUS txt final.zip"/>	Extensões			
Gênero:	<input type="text" value="Científico"/>					
Tipo Textual:	<input type="text" value="Dissertação"/>					
		<input type="button" value="Enviar"/>	<input type="button" value="Limpar"/>			

Fonte: <https://www.etermos.cnptia.embrapa.br/modulo2/modulo2.php>

A segunda etapa possibilita ainda como ferramentas da análise do *corpus*, o total de palavras no *corpus*, total de palavras diferentes e índice de riqueza vocabular. Na etapa 3 de extração automática de termos é realizada o *upload* dos termos previamente estruturados por meio da indexação manual dos textos, conforme ilustra a Figura 10:

Figura 9: Funções da aba “Lista de termos” na terceira etapa do E-termos



Fonte: <https://www.etermos.cnptia.embrapa.br/modulo3/modulo3.php>

A figura 11 mostra a etapa seguinte, a extração de termos, que consiste, em outras palavras, na mineração de tais termos anteriormente inseridos no sistema, e sua verificação nos *corpus*, visando principalmente os índices de frequência simples, como fatores determinantes para as unidades ou conjuntos lexicais se caracterizarem como termos. Observa-se ainda, que o E-termos possibilita a extração Estatística, Linguística e Híbrida, porém, na prática somente a função Estatística estava disponível para uso.

Figura 10: Terceira etapa do E-termos – Extração automática utilizando a função Frequência Simples



Fonte: <https://www.etermos.cnptia.embrapa.br/modulo3/modulo3.php>

Na sequência, a tela apresentada na Figura 12, mostra as opções do extrator automático, onde será selecionado o *corpus* previamente inserido no sistema na Etapa 2, o tamanho do termo que se deseja extrair do corpus, que pode ser de 1 a 7 n-grams, ou seja, de 1 a 7 unidades lexicais ou palavras.

Figura 11: Opções do extrator na Terceira Etapa do E-termos

Fonte: <https://www.etermos.cnptia.embrapa.br/modulo3/modulo3.php#>

As stoplists são as listas de palavras que não são necessariamente caracterizados como termos, mas que devem ser filtradas no momento do processamento da extração de termos, nesta questão, a plataforma dispõe da opção “Padrão do sistema”. A opção valor do corte inferior possibilita ao usuário estabelecer um valor de frequência mínimo para que uma palavra se torne candidata a termo. Para finalizar a fase de identificação que antecede a extração de termos, tem-se a identificação do resultado, ou seja, a nomeação da lista gerada após o processamento.

O sistema apresentou erro ao atender o comando de quatro a sete n-gram, e por este motivo, utilizamos os parâmetros de tamanho do termo (ou n-gram), unigrama (um), bigrama (dois) e trigrama (três) para execução da extração dos *corpus* de teses, dissertações e artigos. No que diz respeito a opção stoplist, a opção disponibilizada para uso foi “Padrão do sistema”.

O valor do corte inferior, ou seja, o quantitativo mínimo de frequência para que uma unidade lexical se candidate a termo, foi estruturada de acordo com a observação da frequência mínima de termos úteis, conforme apresenta a Tabela 6.

Tabela 6: Valor dos cortes de frequência (termos desconsiderados) por gênero

GÊNERO	TAMANHO DO CÓRPUS	CORTE DE FREQUÊNCIA (QUANTITATIVO)
Teses	937.292	100 para unigramas 10 para bigramas e trigramas
Dissertações	6.875.501	100 para unigramas 10 para bigramas 10 trigramas

Artigos científicos	1.669.453	100 para unigramas 10 para bigramas e trigramas
---------------------	-----------	--

Fonte: dados da pesquisa

Inicialmente foram observadas todas as frequências mínimas, e feita a limpeza manual de termos, para se estabelecer um quantitativo válido. Apesar de não haver um consenso na literatura da área sobre os valores de corte, Rijsbergen (1979) afirma que “[...] uma certa arbitrariedade está envolvida na determinação dos pontos de corte, bem como na curva imaginária, os quais são estabelecidos por tentativa de erro.” Nesse sentido, o corte de frequência da pesquisa se baseou nos parâmetros quantitativos de palavras que constituem os *corpus*, e ainda, de acordo com aqueles que mesmo com frequência baixa, no caso 10 (dez), se constituem como termos úteis.

Após a limpeza manual das listas de unigramas, bigramas e trigramas, por gênero textual, geradas pelo E-terms, e da eliminação das palavras que não se constituem necessariamente como um termo. Esse processo teve como resultado o comparativo entre os candidatos a termos por extração estatística e o número final de termos, resultantes da compilação de todos os gêneros textuais para uma visão geral dos dados de processamento de textos na Tabela 7.

Tabela 7: Comparativo entre números de candidatos por extração estatística e número final de termos

N-gram	NÚMERO DE CANDIDATOS DO NSP (Dissertações, teses e artigos)	Nº FINAL DE TERMOS
Unigramas	4.880.851	930
Bigramas	673.644	513
Trigramas	927.961	226
Total	6.482.456	1.669

Fonte: dados da pesquisa

Tais dados mostram que é possível afirmar que, neste caso, quanto maior o número de unidades que compõe o termo, maior o número de candidatos a termos, devido a função “Frequência simples” disponibilizada pelo pacote NSP, integrado ao E-terms. Outro fator que levou à grande redução de termos finais em relação aos candidatos extraídos pelo pacote NSP se deu pela grande quantidade de “sujeira” nos textos processados, a exemplo disso, podemos citar palavras com as acentuações que atrapalham o processo de extração e, conseqüentemente, descarta alguns termos que poderiam vir a ser úteis.

Apesar deste entrave, foi possível realizar a limpeza manual de algum destes termos, e visualizar a frequência de ocorrência, de acordo com o gênero, e o tamanho do n-gram desejado. Os Apêndices de B, C, D, E, F, G, H, I, J e L mostram as unigramas, bigramas e trigramas obtidas no corpus Dissertações, Teses e Artigos científicos, bem como os seus respectivos valores de frequência, a exemplo da Tabela 8.

Tabela 8: Exemplo da lista final de trigramas no corpus Dissertações

Nº	TERMOS	FREQUÊNCIA
1	Escola de música	891
2	Sala de aula	310
4	Música da universidade	224
5	Canto e piano	222
7	História da música	212
Total de trigramas: 7		

Fonte: dados da pesquisa

Em suma, ao final do processo de extração automática de termos obteve-se, em geral, um quantitativo de 930 unigramas, 513 bigramas e 226 trigramas, dos três gêneros textuais analisados, totalizando 1230 termos considerados úteis e representativos para domínio musical. As verificações da coincidência entre as categorias que abrigam os termos, e obtidas por indexação manual e extração automática, estão agrupados na Tabela 9.

Tabela 9: Comparativo entre categorias obtidas entre indexação automática e extração automática

CATEGORIAS DA INDEXAÇÃO MANUAL	CATEGORIAS DA EXTRAÇÃO AUTOMÁTICA
Análise musical	Análise musical
Arranjo	Arranjos
Aspectos emocionais e psicológicos da Música	Compositores
Compositores	Educação Musical
Educação musical	Ensino de Música
Escrita musical	Instrumento musical
Fisiologia vocal	Música vocal
Função social da Música	Nota musical
Gênero	Partitura
Grupo rítmico	Performance

História da Música	Prática musical
Instrumento musical	Processo composicional
Legislação	Processos de estúdio
Música e cultura	Uso da música
Música vocal	
Nota musical	
Performance	
Prática (campo) profissional	
Prática Interpretativa	
Processo composicional	
Processo de estúdio	
Produção vocal	
Psicologia cognitiva musical	
Recurso computacional	
Tempo musical	
Teoria musical	
Textura musical	
Uso (execução) da Música	
Vibração de instrumento	

Fonte: dados da pesquisa.

Dentre as categorias estruturadas, que abrigam os termos provenientes da extração automática, estão: análise musical, arranjos, compositores, educação musical, ensino de Música, instrumentos musicais, música vocal, nota musical, partitura, performance, prática musical, processos composicionais, processos de estúdio e uso da música.

Das 29 categorias manuais estruturadas, somente três indicaram baixa frequência de termos pertencentes na extração automática, a saber: fisiologia musical, função social da Música, grupos rítmicos, história da Música, legislação e textura musical.

Tais considerações mostram que, apesar de o método estatístico gerar ruídos, ou seja, palavras que não possuem valor terminológico, este método é de extrema importância para fins de indexação, tradução, construção de tesouros, entre outras ferramentas de representação e recuperação da informação por proporcionar automatizar a identificação e seleção de unidades lexicais de um *corpus*, ao proporcionar rapidez na construção de Terminologias.

Os subsídios apontados a partir desta análise mostram que a extração automática de termos baseada em frequência estatística, facilitada pelo ambiente E-terms, permite o aprimoramento das técnicas e reforça a precisão no processo de Indexação. Espera-se ter demonstrado o percurso para estruturação de um vocabulário com os subsídios do Processamento da Linguagem Natural para construção de vocabulários.

CONCLUSÃO

A pesquisa se deteve em abordar a prática da indexação e representação da informação para estruturação do vocabulário do domínio musical no período de uma década, de 2003 a 2013, com a finalidade de analisar os léxicos produzidos nesta área e relacioná-los com outros conceitos, por meio da categorização.

Dentre os objetivos da análise prático-conceitual da indexação, destacam-se como resultados os instrumentos de representação da informação musical, como catálogos, índices e tesouros com o objetivo de servir como recurso e/ou ferramenta informacional auxiliar para pesquisadores e interessados em geral na busca e recuperação da informação musical. Porém, percebe-se que os instrumentos existentes ainda são escassos e falhos no que diz respeito ao conteúdo dos documentos, para o suprimento da necessidade informacional de seus usuários.

Em se tratando da área de Música, não foi localizada obra lexicográfica e/ou terminológica sobre o domínio, com contribuições dos aportes automáticos, em especial de PLN, para sua estruturação, o que evidencia a contribuição desta pesquisa ao tentar minimizar a carência de informações sobre o controle do léxico utilizado pela área musical.

Observa-se ainda a importância da categorização para a identificação de assuntos de um domínio de especialidade, devido à possibilidade de abordagem facetada de conteúdos que permite a visualização de uma área do conhecimento como um todo sistematizado, e viabilizando relações com outros conceitos.

O resultado deste processo foi a atribuição de 29 categorias, sendo que destas, as que tiveram mais termos agregados na Indexação Manual foram Prática interpretativa (121 termos), Processo composicional (100 termos) e Educação Musical (97 termos). A extração automática por meio do E-terms, por sua vez, mostrou-se eficaz ao apresentar um quantitativo de 922 itens lexicais, entre unigramas, bigramas e trigramas que podem aprimorar a construção de vocabulários fundamentados em PLN.

Com relação à análise comparativa entre os processos manual e automático de extração de termos, a convergência entre estes dois métodos consiste na subjetividade humana para seleção e correção dos termos encontrados, porém, vale ressaltar que a intersecção entre as categorias obtidas por Indexação manual e as categorias geradas por extração automática alcançaram índices de frequências diferentes durante o processo. Isso significa que, nem todos os termos elencados no método manual foram extraídos automaticamente.

Nesse sentido, a escolha pelo método estatístico se deu pelo fato de a frequência apresentar um quantitativo maior de descritores significativos, que proporcionam uma representação e recuperação mais precisa de termos, além da rapidez na extração diante de grandes volumes de textos.

A principal problemática encontrada durante o processo de extração automática de candidatos a termos consiste na disponibilidade dos softwares (híbridos, linguísticos ou estatísticos) de forma gratuita, haja vista que tais softwares citados durante o referencial teórico encontram-se em teste nos Programas de Pós-Graduação voltados para a Computação/Inteligência Artificial e Linguística Computacional e, por este motivo, ainda não foram disponibilizados para a comunidade acadêmica.

Outra questão envolve a interface com o usuário, a exemplo do *GATE* e Pacote NSP que foram idealizados inicialmente para a pesquisa, porém, os seus formatos via linha de código, ou seja, onde o software para seu funcionamento necessita de códigos para programação de computadores, ocasionaram entraves não somente para este estudo.

Vale retomar algumas considerações a respeito da extração automática, e suas abordagens observadas no desenvolvimento da pesquisa: na Terminologia, a extração automática corresponde à aquisição de um produto terminológico que representa os léxicos, a exemplo de dicionários, índices ou glossários; enquanto a Computação a entende como abordagem automática de reconhecimento e extração de termos de uma especialidade, geralmente realizada por meio das ferramentas de PLN.

A extração de candidatos a termos se mostra proveitoso também para outras utilidades, tais como: recuperação da informação, construção de ontologias, sumarização automática, tradutores, alinhamento de textos, corretores gramaticais e PLN. Com a utilidade da extração automática minimiza-se a subjetividade do indexador por meio do uso dos subsídios da Inteligência Artificial que tende a aproximação do raciocínio homem-máquina para solução de questões relacionadas à linguagem.

Apesar de todo o vasto subsídio teórico disponível pela Linguística Computacional, em especial pelo PLN, em termos práticos ainda existem lacunas na prática no Brasil, que se encontra ainda em fase de desenvolvimento e aprimoramento de softwares de forma a conduzi-los a efetivação da interpretação da linguagem humana. O que nos leva a crer que ainda existe um longo percurso a ser feito para abranger todos os domínios do conhecimento.

A literatura revela que grande parte das pesquisas em PLN concentra-se em áreas e subdomínios: manipulação de bases de dados, sistemas tutores, sistemas de processamento de textos científicos, sistemas especializados, tradução automática, sistemas acadêmicos, geração de resumos e extração da informação.

Assim com base nos pressupostos teórico-metodológicos dos estudos do Léxico, Linguística de *Corpus* e PLN estudados e empregados, foi possível a elaboração do vocabulário de Música, a ser estendido e aprimorado. Embora se tenham atingido os propósitos aqui pretendidos, é importante considerar os passos previstos para continuidade da pesquisa, como:

- Extração de termos baseada em técnicas estatísticas, para execução de outras ferramentas;
- Criação e processamento de um novo corpus que envolva outras tipologias de produções técnico-científicas do domínio musical;
- Construção de Ontologias do domínio musical.

O PLN mostrou-se como uma ferramenta eficaz para processamento de grandes volumes de dados, com muito a contribuir no que diz respeito à redução do tempo de desempenho de tarefas de mineração de textos e ao possibilitar a identificação dos termos mais utilizados para representação de um domínio. Apesar destas contribuições, destaca-se que a intervenção humana ainda é necessária para a limpeza dos materiais obtidos e para a validação dos resultados.

Espera-se ter proposto uma metodologia para elaboração de novas representações de domínios, com o intuito de aprimorar as técnicas e ferramentas de representação da informação utilizadas na Ciência da Informação.

Diante de tais considerações chega-se a conclusão de que o vocabulário de Música é uma ferramenta útil e enriquecedora no sentido da apresentação tanto para os próprios músicos, ao facilitar a aproximação e a conseqüente recuperação dos termos utilizados na linguagem natural para a linguagem artificial no momento de busca em sistemas, por meio

do PLN, quanto para os linguistas e bibliotecários, ao fornecer subsídios que contribuem para o controle dos léxicos.

REFERÊNCIAS

ALUÍSIO, S. M.; ALMEIDA, G. M. de B. O que é e como se constrói um corpus: lições aprendidas na compilação de vários corpora para pesquisa lingüística. **Calidoscópico** (UNISINOS), vol. 4, n. 3, p. 155-177, set./dez. 2006. Disponível em: <http://www.unisinos.br/publicacoes_cientificas/images/stories/pdfs_calidoscopio/vol4n3/art04_aluisio.pdf>. Acesso em 25 mar. 2014.

AMERICAN NATIONAL STANDARDS INSTITUTE. National Information Standards Organization. **ANSI/NISO Z39.19-2003**: guidelines for the construction, format and management of monolingual thesauri. Disponível em: <<http://www.niso.org/kst/reports/standards>>. Acesso em: 20 ago. 2013.

ANDRADE, M. M. Lexicologia, Terminologia: definições, finalidades, conceitos operacionais. In: OLIVEIRA, A. M. P. P.; ISQUERDO, A. N. (Org.). **As ciências do léxico: lexicologia, lexicografia, terminologia**. 2. ed. Campo Grande, MS: UFMS, 2001.

ANTONIO, Irati. **Informação e Música no Brasil**: memória, história e poder. São Paulo, 1994. 285 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo.

ARTÊNCIO, Luciane Maria. **Princípios de categorização nas linguagens documentárias**. 2007. 129 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2007.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12. 676**: Métodos para análise de documentos – determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro, 1992.

BAGOT, R. E. **Extracció de terminologia: elements per a la construcció d'un SEACUSE**. 1999. Tese (Doutorado em Linguística Aplicada) – Institut Universitari de Linguística Aplicada, Universitat Pompeu Fabra, Barcelona, 1999.

BARBOSA, Maria Aparecida. Lexicologia, lexicografia, terminologia e terminografia: identidade científica, objeto, métodos, campos de atuação. In: Simpósio Latino-Americano de Terminologia, 2., 1990. Curitiba. **Anais...** Curitiba: IBICT, 1990.

BERBER SARDINHA, T. Linguística de corpus: histórico e problemática. **DELTA**, São Paulo, v. 16, n. 2, 2000.

BERBER SARDINHA, T. Tamanho de corpus. **The Specialist**, São Paulo, v. 23, n. 2, p. 103-122, 2003.

BECKER, David. Automated Language Processing, **Annual Review of Information Science and Technology**, v. 16, p. 113-138, 1981.

BIBER, D.; CONRAD, S.; REPPEN, R. **Corpus linguistics: investigating language structure and use**. Cambridge: Cambridge University Press, 1998.

BOBROW, D. G.; FRASER, J. B.; QUILLIAN, M. R. Automated Language Processing, **Annual Review of Information Science and Technology**, v. 2, p. 161-186, 1967.

BORKO, H. Information Science: what is it? **American Documentation**, v. 19, n. 1, p. 3-5, jan. 1968.

BRIGGS, Asa; BURKE, Peter. Uma história social da mídia: de Gutemberg à Internet. Rio de Janeiro: Jorge Zahar Ed., 2004.

BUSH, Vanevar. As we may thing. [LOCAL]: *Atlantic Monthly*, p. 101-108, 1945.

CABRÉ, M. T. **La terminologia**: teoria, metodologia, aplicaciones. Barcelona: Antártida; Empúries, 1993.

CAMPOS, Maria Luiza de Almeida; GOMES, Hagar Espanha. Organização de domínios de conhecimento e os princípios rangenathianos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 8, n. 2, p. 150-163, jul./dez. 2003.

CAMPOS, Maria Luiza de Almeida. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ciência da informação**, Rio de Janeiro, v. 33, n. 1, p. 22-32, jan./abril 2004.

CANDIDO JUNIOR, Arnaldo. **Criação de um ambiente para o processamento de corpus de português histórico**. 2008. 142 f. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2008.

CESARINO, Maria Augusta da Nóbrega. Sistema de recuperação da informação. **Revista da Escola da UFMG**. Belo Horizonte, v. 14, n. 2, p. 157-168, set. 1985.

CINTRA, Anna Maria Marques et al. **Para entender as linguagens documentárias**. 2. ed. rev. e ampl. São Paulo: Polis, 2002.

CHOWDHURY, Gobinda C. Natural Language Processing, Annual Review of Information Science and Technology, v. 37, p. 51-89, 2003.

CURRÁS, Emília. **Tesouro, linguagens terminológicas**. Brasília: IBICT, 1995.

DAHLBERG, Ingetraut. Teoria do conceito. **Ciência da Informação**, Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978.

DALMOLIN, Luiz Claudio Duarte. **Extrator de termos para criação de mapas conceituais**. 2010. 118 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Santa Catarina, Florianópolis, 2010.

DEMO, Pedro. Pesquisa e construção de conhecimento. Rio de Janeiro: Tempo Brasileiro, 1996.

DIAS, Eduardo Wense; NAVES, Madalena Martins Lopes. **Análise de assunto: teoria e prática**. Brasília: Thesaurus, 2007.

DODEBEI, Vera Lucia Doyle. **Tesouro: linguagem de representação da memória documentária**. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002.

DUDEQUE, Norton. Sobre harmonia de Arnold Schoenberg. **Per Musi**. Belo Horizonte, v. 9, jan-jun./2004. p. 114-123.

E-TERMOS, Ambiente Colaborativo Web de Gestão Terminológica. 2014. Disponível em: <<https://www.etermos.cnptia.embrapa.br/>>. Acesso em: 10 jul. 2014.

FERREIRA, Sueli. **O ensino das artes: construindo caminhos**. Campinas: Papirus, 2001.

FUJITA, Mariângela Spotti Lopes. Modelos de categorização para construção de tesouros: metodologia de ensino. In: VERA REGINA CESARI BOCCATO; LUCIANA DE SOUZA

GRACIOSO (Orgs.). **Estudos de linguagem em Ciência da Informação**. Campinas: Editora Alínea, 2011. p. 35-68.

GIRARDI, Rosario; DRUMOND, Lucas Rêgo. Infonorma: um sistema de recomendação baseado em tecnologia da Web Semântica. **Infocomp (UFLA)**, v. 5, p. 93-102, 2006.

GRANJA, C. E. de S. C. **Musicalizando a escola: música, conhecimento e educação**. São Paulo: Escrituras Ed., 2006.

GUINCHAT, Claire; MENOU, Michel. A indexação. In: **Introdução geral às ciências e técnicas da informação e documentação**. 2. ed. Brasília: IBCT, 1994.

HARMON, G. *On the evolution in information Science*. **Journal of the American Society for Information Science**, v. 22, n. 4, p. 235-241, july/ aug. 1971.

HUMMES, J. **As funções do ensino de música, sob a ótica da direção escolar: um estudo nas escolas de Montenegro/RS**. Dissertação (Mestrado em Educação Musical) – Programa de Pós-Graduação em Música, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.

ISO 704. Terminology work - principles and methods. 2.ed. Genève: International Standard Organization, 2000.

KOBASHI, Nair Yumiko; FRANCELIN, Marivalde Moacir. Conceitos, categorias e organização do conhecimento. **Informação e informação**, v. 16, n. 2, p. 1-24, jan./jul. 2011. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/10390/9281>>. Acesso em 21 mar. 2014.

LÁCIO-WEB. 2014. Disponível em: <<http://www.nilc.icmc.usp.br/lacioweb/>>. Acesso em: 05. mar. 2014.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2. ed. Brasília: Briquet de Lemos, 2004.

LANCASTER, F. W. *Information retrieval systems: characteristics, testing and evaluation*. New York: John Wiley & Sons, 1968.

LARA, Marilda Lopez Ginez de. Linguagem documentária e terminologia. **Transinformação**, Campinas, v. 16, n. 3, p. 231-240, set./dez. 2004. Disponível em: <<http://revistas.puc-Campinas.edu.br/transinfo/viewarticle.php?id=72>>. Acesso em: 20 ago, 2013.

LINGUATECA. 2014. Disponível em: <<http://www.linguateca.pt/>>. Acesso em 14 abr. 2014.

LIU, M. **Hypermedia assisted instruction and second language learning: a semantic-network-based approach**. [S.l.]: The Haworth Press, Inc., 1994. p. 293-312.

LOPES, Lucelene; VIEIRA, Renata. Processamento de linguagem natural e o tratamento computacional de linguagens científicas. In: PERNA, Cristina Lopes; DELGADO, Heloísa Koch; FINATTO, Maria José (Orgs.). **Linguagens especializadas em corpora: modos de dizer e interfaces de pesquisa [recurso eletrônico]**. Porto Alegre: EDIPUCRS, 2010.

MAI, J. E. Deconstructing the indexing process. **Advances in Librarianship**, 23, 2000, p. 269-298.

MAI, J. E. Semiotics and indexing: na analysis of the subject indexing process. **Journal of Documentation**, 57, 2001, 591-622.

McDONALD, Carlton; YAZDANI, Masoud. **Prolog programming: a tutorial introduction**. Oxford: Blackwell Scientific Publications, 1990.

MEDEIROS, Marisa Bräsher Basílio. Terminologia brasileira em ciência da informação. **Ciência da Informação**, v. 15, n. 2, p. 135-142, jul./dez. 1986.

MELLO, Maria Ignez C. **Música popular brasileira e estudos culturais**. 1997. 37 f. Trabalho de conclusão de curso (Especialização) – Universidade Federal de Santa Catarina, Florianópolis, 1997.

MERRIAM, A. P. The anthropology of music. Evanston: Northwestern University Press, 1964.

MOOERS, C. N. Zatocoding applied to mechanical organization of knowledge. **American Documentation**, v. 2, p. 20-32, 1951.

MOLLICA, Maria Cecilia (Org.). Linguística e Ciência da Informação: diálogos possíveis. Curitiba: Editora Appris, 2012.

MONTGOMERY, C. A. Linguistics and Information Science. *Journal of the American Society for Information Science*, v. 23, n. 3, p. 195-219, maio-jun. 1972.

MORAES, J. S.; CRISTIANINI, G. M. S. Revisão de vocabulário controlado e critérios para a seleção de literatura: o caso da área de Estatística e Probabilidade do ICMC/USP. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 16., 2010, Rio de Janeiro. **Anais...** Rio de Janeiro: UFRJ, 2010. Disponível em: <http://www.gapcongressos.com.br/eventos/z0070/trabalhos/final_296.pdf>. Acesso em: 17 abr. 2012.

NANTES, L. M. **Desenvolvimento de um sistema baseado em linguagem natural para consultas em banco de dados na Web**. 63 p. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade do Oeste Paulista, Presidente Prudente, 2008. Disponível em: <http://fipp.unoeste.br/~chico/FIPP/projetos/projeto2008/Monografia_Nantes_2008.pdf>. Acesso em: 20 ago. 2013.

NUNES, M. G. V.; DIAS-DA-SILVA, B. C.; RINO, L. H. M.; OLIVEIRA JR., O. N.; MARTINS, R. T.; MONTILHA, G. **Introdução ao processamento das línguas naturais**. Notas Didáticas do ICMC, n. 38. São Carlos/SP, 1999. p. 91.

OLIVEIRA NETO, J. M.; TONIN, S. D.; PRIETCH, S. S. **Processamento de linguagem natural e suas aplicações computacionais**. 2010. Disponível em: <<http://www.inpa.gov.br/erin2010/Artigo/Artigo9.pdf>>. Acesso em: 20 ago. 2013.

OLIVEIRA, Marlene de (Org.). **Ciência da informação e biblioteconomia: novos conteúdos e espaços de atuação**. 2. ed. Belo Horizonte: Editora UFMG, 2011.

PAZIENZA, M. T. et. al. *Terminology extraction: na analysis of linguistic and statistical approaches*. **Studies in fuzziness and soft computing**, v. 185, p. 255-280, 2005.

PINHEIRO, L. V. R. **A ciência da informação entre a sombra e a luz: domínio epistemológico e campo interdisciplinar**. 1997. 269 f. Tese (Doutorado em Comunicação) – ECO, Universidade Federal do Rio de Janeiro, Rio de Janeiro/Instituto Brasileiro de Informação em Ciência e Tecnologia, 1997.

PINTO MOLINA, M. *Análisis documental: fundamentos y procedimientos*. 2. ed. rev. e ampl. Madri: Eudema, 1993.

PRAXEDES FILHO, Pedro Henrique Lima. **Linguística sistêmico-funcional: linguística teórica ou linguística aplicada**. In: Seminário Interdisciplinar das Ciências da Linguagem no Cariri, 1., 2012. p. 1-52. Disponível em: <http://urca.br/sic/i/arquivos/slide_pedro_henrique.PDF>. Acesso em: 10 set. 2013.

QUEIROZ, Gregório José Pereira de. **A música compõe o homem, o homem compõe a música**. São Paulo: Cultrix, 2000.

RANGANATHAN, S. R. **Prolegomena to Library classification**. Bombay: Asia Publishing House, 1967. 640 p.

RIJSBERGEN, V. C. J. *Information retrieval*. 2. ed. Glasgow: Dept. of Computer Science, University of Glasgow, 1979.

ROBREDO, J. **Documentação de hoje e de amanhã: uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas**. 4. ed. ver. e ampl. Brasília: Edição de autor, 2005.

RODRIGUES FILHO, Ilson Wilmar. *Processamento de linguagem natural*. 2004. Disponível em: <<http://www.inf.ufsc.br/~ilson/slides.ppt>>. Acesso em: 15 mar. 2014.

SANTINI, Rose Marie. Recuperação da informação de música e a Ciência da Informação: tendências e desafios de pesquisa. In: ENANCIB, 8, 2007, Salvador. **Anais...** Salvador: UFBA, 2007. p. 1-14.

SARACEVIC, T. Information Science: origin, evolution and relation. In: VAKKARI, P.; CRONIN, B. (Ed.). **Conceptions of library and information science**. Proceedings of the international conference for the celebration of 20th anniversary of the Department of Information Studies, University of Tampere, Finland, 26-28, 1991. London; Los Angeles: Taylor Graham, 1992, p. 5-27.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n. 1, p. 41-62, jan./jun. 1996.

SILVA, Renato Rocha; LIMA, Sérgio Muinhos Barroso. Consultas em bancos de dados utilizando linguagem natural. **Revista Eletrônica da Faculdade Metodista Granbery**, Juiz de Fora, v. 7, n. 2, ago/dez. 2007. Disponível em: <<http://re.granbery.edu.br/artigos/MjQ0.pdf>>. Acesso em: 30 ago. 2013.

SILVA, Bento Carlos Dias da. O estudo linguístico-computacional da linguagem. **Letras de Hoje**, v. 41, n. 2, p. 103-138m abr./jun. 2006. Disponível em: <<http://revistaseletronicas.pucrs.br/ojs/index.php/fale/article/view/597/428>>. Acesso em 25 ago. 2013.

SILVEIRA, Filipe Pereira da. **Integração de ferramentas para compilação e exploração de corpora**. 2008. 99 f. Dissertação (Mestrado em Ciência da Computação), Faculdade de Informática, PUCRS, 2008.

SINCLAIR, J. Corpus, concordance, collocation. Oxford: Oxford University Press, 1991.

TEIXEIRA, Rosana de Barros Silva e. **Termos de (Onco)mastologia: uma abordagem mediada por corpus**. 2010. 392 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2010.

TRISTÃO, Ana Maria Delazari; FACHIN, Gleisy Regina Bóries; ALARCON, Orestes Estevam. Sistema de classificação facetada e tesouros: instrumentos para organização do conhecimento. **Ciência da Informação**, Brasília, v. 33, n. 2, p. 161-171, maio/ago. 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33/n2/a17v33n2.pdf>>. Acesso em 20 ago. 2013.

TURAZZA, J. S. **Léxico e criatividade**. São Paulo: Plêiade, 1996.

VIEIRA, R.; LIMA, V. L. S. Linguística computacional: princípios e aplicações. In: IX Escola de Informática da SBC-Sul. Luciana Nedel (Ed.). Passo Fundo, Maringá, São José. SBC-Sul, 2001.

WANDERLEY, M. Linguagem documentária: acesso à informação. **Ciência da Informação**, Rio de Janeiro, v. 2, n. 2, p. 175-217, 1973.

WALKER, Donald E. Automated Language Processing, **Annual Review of Information Science and Technology**, v. 8, p. 69-119, 1973.

WARNER, A. J. Natural Language Processing, **Annual Review of Information Science and Technology**, v. 22, p. 79-108, 1987.



DESVENDANDO A DIMENSÃO COGNITIVA DA LEITURA NO PROCESSO DE INDEXAÇÃO

Tatiane da Silva Mendonca

INTRODUÇÃO

O objetivo desse capítulo é conhecer e analisar os aspectos cognitivos que auxiliam o bibliotecário no processo de leitura realizado na indexação. Para tanto, faz-se importante contemplar algumas ressalvas conceituais, no intuito de estabelecer tal relação cognitiva.

Para Neves (2012, p. 40), por exemplo, indexação “é um processo subjetivo que depende, em grande parte, do desempenho do indexador.” Ou seja, é uma atividade que exige do bibliotecário indexador aptidões e conhecimentos prévios sobre determinado assunto, que são obtidos a partir da observação ou percepção de mundo, bem como por meio dos registros que estão em sua mente (memória ou pensamento) visando a representação adequada da informação.

Boccatto (2011, p. 17, grifo nosso), por sua vez, afirma que,

[...] a análise e identificação do conteúdo [as quais caracterizam a primeira etapa do processo de indexação] são realizadas por meio da cognição, por isso ela não é um processo sistemático de etapas consecutivas, mas é um processamento mental de informações, que depende de conhecimento prévio e é composta de habilidades operatórias de pensamento.

Neste contexto, pode-se dizer que o processo de indexação envolve tanto o uso da cognição do indexador, somado à sua percepção, memória e representação mental do conhecimento e aos conhecimentos prévios, intrínsecos aos indexadores (que sofrem alterações com base no ambiente em que está inserido), quanto aos conhecimentos específicos (estratégias de leitura) que propiciarão um bom desempenho desta tarefa.

Outro ponto a se considerar refere-se ao contexto sociocognitivo do bibliotecário indexador que é um fator imprescindível para o desenvolvimento do processo de indexação - o qual dá-se, inicialmente, pela leitura documental, para que após a análise do assunto o indexador possa representar a informação contida no documento, conforme as necessidades dos usuários. Em outras palavras, pode-se entender que a representação da informação necessária para a construção de sistemas de recuperação irá depender do contexto em que se dá o processo, ou seja, o indexador sempre buscará atender às necessidades e expectativas dos usuários o que dependerá do tipo de biblioteca (pública, escolar, universitária etc.) em que se encontra este profissional e usuário. Assim sendo, compreende-se que o contexto sociocognitivo do bibliotecário indexador

[...] deve ser visto como um elemento norteador no desempenho do bibliotecário indexador, pois estabelece e delimita as funções desse profissional dentro da biblioteca, de modo que ele possa exercer suas atividades de acordo com o interesse dos usuários que utilizam o sistema de recuperação da informação. (BOCCATO, 2012, p. 80).

Dentre os aspectos que cercam o campo da indexação, destaca-se a análise conceitual/documentária ou de assunto que refere-se ao “conjunto de procedimentos efetuados com o fim de expressar o conteúdo de documentos, sob formas destinadas a facilitar a recuperação da informação.” (CUNHA, 1987, p. 39). Ou, como sugere Rubi (2009, p. 82), trata-se de uma atividade “[...] realizada por meio da leitura documentária feita pelo indexador, que procura compreender de maneira geral o documento para identificar e selecionar os termos que o representarão para efeito de recuperação.”

Neste contexto, compreende-se que a análise de assunto envolve antes de mais nada a leitura do texto, que se configura como uma leitura interativa que relaciona o texto, o leitor e o seu contexto, cuja finalidade é representar aquilo que realmente caracteriza o conteúdo do documento. Considerando que esta prática vai além da decodificação de sinais visuais e que envolve a qualidade do texto, o conhecimento prévio e as estratégias de leitura adotadas pelo bibliotecário, então reconhece-se que, na verdade, trata-se de um processo subjetivo que compreende os “aspectos cognitivos relacionados com conhecimentos armazenados na memória do leitor [bibliotecário] e comportamentos específicos durante a leitura.” (CINTRA, 1987, p. 29). Assim, entende-se que a leitura é uma atividade extremamente complexa, que associa estruturas linguísticas (a forma como os textos são organizados) e processos cognitivos - que correspondem àquilo que o sujeito cognoscente assimila, compreende e representa mentalmente como sendo verdadeiro dos estímulos externos a ele apresentados.

Assim, dentre as áreas que abordam esta temática, a saber: a Linguística, a Semiótica e a Lógica, vai se tratar com mais afinco dos aspectos que integram a Psicologia Cognitiva, no tocante à leitura como um processo cognitivo complexo, pois revela ser o indexador-leitor um sujeito cognoscente que se utiliza de aspectos cognitivos para a prática de análise conceitual.

Neste sentido, apresenta-se a seguir os principais fatores que interferem/auxiliam diretamente no processo de indexação, tendo por instrumento de pesquisa os questionários aplicados aos bibliotecários que são responsáveis pela representação temática da informação nas bibliotecas universitárias de São Luís e o uso de uma dinâmica – “*Indexando textos*”, cuja finalidade era verificar a prática das bibliotecárias no processo de análise de assunto, ou seja, como realizam a leitura do documento até chegar à sua tradução (passagem de uma linguagem natural para uma linguagem documentária), visando à recuperação pelos usuários.

DIMENSÃO CONCEITUAL DO PROCESSO DE INDEXAÇÃO: o olhar do bibliotecário

O bibliotecário indexador tem importância singular no processo de representação temática da informação e, como tal, este profissional deve conhecer com propriedade as etapas que compreendem este processo, são elas: análise de assunto/conceitual e tradução, uma vez que deste processo resultará e/ou implicará diretamente no processo de recuperação da informação pelos usuários. Neste sentido, no primeiro bloco de perguntas dos questionários objetivou-se saber o grau de conhecimento e domínio das bibliotecárias com relação ao processo de indexação. Perguntou-se, então, sobre o entendimento acerca do conceito, dos instrumentos e métodos utilizados, da existência de uma política de tratamento documental, da etapa considerada mais difícil neste processo e das competências que julgavam ser imprescindíveis para o processo de indexação.

Em geral, as bibliotecárias questionadas compreendem que a indexação é o processo pelo qual se realiza uma leitura técnica, a fim de extrair os termos/descriptores/palavras-chave que mais representam/identificam o conteúdo de um documento, de forma a propiciar sua organização e recuperação pelos usuários, conforme observa-se nas descrições abaixo:

- é a ordenação, estruturação das palavras extraídas de um documento o qual permite a identificação do tema ou assunto na recuperação da informação. (R/CEST);
- significa a descrição e identificação do documento. (JO/UFMA);

- ferramenta de recuperação da informação que identifica as condições de acesso na biblioteca (palavras-chave). (J/UFMA);
- [...] consiste na leitura técnica que determina termos ou palavras do assunto chave do documento. (B/UEMA);
- é o ato de identificar um documento e/ou descrevê-lo de acordo com o seu tema (análise dos aspectos temáticos da informação documental). (G/UEMA).

Neste contexto, compreende-se que o processo de indexação equivale ao “[...] ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p. 2).

No que tange aos instrumentos e métodos utilizados pelas bibliotecárias durante o processo de indexação, destacam-se: vocabulário livre - no caso da biblioteca da UEMA, que caracteriza-se “pela ausência de um controle do vocabulário usado para a indexação” (NOVELLINO, 1996, p. 39); Anglo American Cataloging Rules (AACR2); Tabela Cutter; Buscadores (Google); Catálogos ou índices; Tesouro e a Classificação Decimal Universal - que, apesar de não ser mencionada nas respostas, foi utilizada por uma das bibliotecária durante a dinâmica, como forma de verificar se os termos extraídos condiziam com os descritos na tabela de classificação. Conforme a literatura pertinente à questão dos métodos e instrumentos utilizados no processo de indexação, o que mais se destaca é o uso de linguagem documentária, principalmente, em decorrência da segunda etapa deste processo, a tradução. De acordo com Novellino (1996, p. 38) a linguagem documentária enquanto “[...] instrumento de padronização da indexação, [...] visa garantir que indexadores de um mesmo sistema ou sistemas afins usem os mesmos conceitos para representar documentos semelhantes.”

Além da linguagem documentária, existem as listas de cabeçalhos de assunto que também propiciam a padronização dos termos usados no processo de indexação e o tesouro que funciona como “[...] instrumento facilitador da comunicação dentro do sistema, padronizando as linguagens de indexação e de recuperação, a partir da terminologia da área representada.” (NOVELLINO, 1996, p. 40). Dentre as vantagens do uso de tesouros como instrumentos, está o fato de permitir ao usuário observar, detalhadamente, a organização de uma determinada área do conhecimento, e no mais, “[...] indexadores e usuários estão mais familiarizados com os termos expressos em linguagem natural de um tesouro do que com as notações de um sistema de classificação.” (NOVELLINO, 1996, p.

41), o que facilita, consideravelmente, a recuperação da informação. O que também, auxilia na recuperação precisa da informação é a presença de uma política de tratamento documental, visto que garante a padronização e uniformidade do processo de indexação o que refletirá no processo de representação, organização e recuperação da informação. Perguntou-se então às bibliotecárias sobre a existência de uma política de indexação na sua instituição e a resposta foi positiva, ou seja, todas confirmaram a existência de uma política, o que mostra que as bibliotecas têm se preocupado com a qualidade dos seus produtos e serviços meios e fins, uma vez que o estabelecimento de uma política documental garante a eficácia na recuperação da informação, que é objetivo fim do processo de tratamento da informação. No entanto, vale destacar que, em atenção à literatura, ficou claro que a existência formalizada e documentada de uma política de indexação na prática não existe, o que existem são práticas cotidianas e repetitivas que acabam passando-se por política, daí compreende-se que as afirmativas das bibliotecárias referem-se justamente à rotina de indexação e não propriamente a uma política.

Outro ponto importante foi saber que para a maioria das bibliotecárias (com exceção de uma bibliotecária que considerou a tradução como etapa mais criteriosa), a análise de assunto é a etapa mais difícil do processo de indexação, visto que exige que o bibliotecário desenvolva uma leitura minuciosa e cautelosa, a qual envolve uma série de fatores, que vão desde o conhecimento geral acerca do documento até a forma como este está estruturado, textualmente falando.

Segundo Fujita e Rubi (2006, p. 2),

As dificuldades existem porque a leitura é um processo de interação com o texto escrito visando a sua compreensão e isso significa um processo de cognição. O processo de análise de assunto para a indexação, dessa forma, envolve a compreensão do texto mediante processos cognitivos, realizados com base em esquemas mentais.

Neste contexto, compreende-se que a análise de assunto, precisamente, a leitura documentária é uma atividade complexa que envolve, não só conhecimentos de mundo ou prévios, mas linguísticos, lógicos e cognitivos, daí a dificuldade revelada pelas bibliotecárias ao considerar tal etapa como sendo a mais criteriosa e portanto a mais difícil.

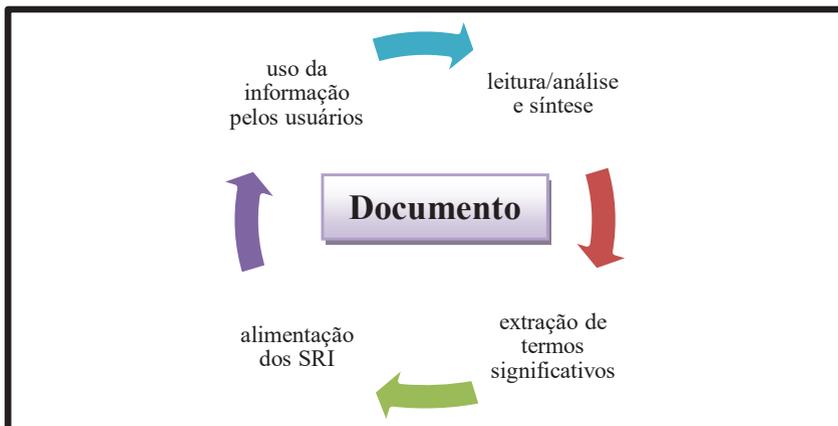
Sabendo da importância do processo de indexação para a recuperação precisa da informação, perguntou-se quais seriam as competências que os bibliotecários indexadores deveriam apresentar para a realização da indexação? Em geral, as bibliotecárias destacaram: a atenção, o gosto pela leitura, conhecimento do assunto, noções de indexação (conhecer as etapas que envolvem o processo) e ser sensível à leitura técnica do

documento, de forma a identificar e selecionar os termos/conceitos e os representar por meio de linguagens documentárias. O processo de indexação exige que o bibliotecário tenha responsabilidade e principalmente, aprofundamento teórico acerca dos temas a serem indexados, uma vez que a extração de qualquer termo, sem obedecer as necessidades dos usuários e sem representar fielmente o documento pode causar prejuízo na hora de organizar e recuperar este documento. Neste sentido, Cunha (2000, p. 71) salienta que inúmeros fatores têm contribuído para se repensar tal prática, dentre eles:

A substituição dos paradigmas tradicionais das profissões da informação como consequência do impacto das novas tecnologias sobre o processamento, a transmissão, a organização e o acesso à informação, a ubiquidade da informação disponível e seu acesso virtual - tudo contribui para repensar competências, habilidades e estratégias de formação para um exercício profissional satisfatório.

Ou seja, além das práticas já exigidas dos bibliotecários durante o processo de indexação, há também a necessidade de este profissional dominar as novas tecnologias, visto que se fazem cada vez mais presentes em seu dia a dia e refletem diretamente nos processos que delas fazem uso, como o processamento da informação (ver Figura 1).

Figura 1 – Processamento da informação



Fonte: Autoria própria.

Conforme a figura 1, compreende-se que o processamento da informação envolve a leitura/análise e síntese do documento; extração de termos significativos; alimentação dos sistemas de recuperação da informação e uso da informação pelos usuários. Trata-se, portanto, de um ciclo documental e, para que o produto final seja eficiente, é necessário evitar quaisquer interferência durante este ciclo.

DIMENSÃO COGNITIVA DO PROCESSO DE LEITURA NA INDEXAÇÃO: percepção, memória e representação mental do conhecimento

O segundo bloco de perguntas envolve questões ligadas às estratégias de leitura desenvolvidas pelas bibliotecárias, bem como os aspectos que interligam o processo de indexação, no tocante à análise de assunto (leitura documental) e a Psicologia Cognitiva, no que tange ao uso da percepção, da memória e da representação mental do conhecimento, o que resultará na representação, organização e recuperação precisa e eficiente dos documentos/informação pelos usuários.

Nesta perspectiva, buscou-se por meio de perguntas direcionadas aos bibliotecários, saber quais estratégias de leitura, ou melhor, quais partes do documento, consideravam para realizar a análise documental, com fins de representação e recuperação de informação, levando em consideração que a leitura documental é a etapa primordial para o processo de indexação. As bibliotecárias leem, em geral, a folha de rosto, a orelha do livro, o resumo, o sumário, a introdução, a conclusão e, caso necessário, realiza a leitura da parte introdutória de cada capítulo etc. Comparando-se com a literatura, as bibliotecárias seguem as dicas presentes na maioria dos manuais e livros especializados, pois de acordo com Dias e Naves (2007, p. 29, grifo nosso), “As partes mais comumente citadas nos manuais de catalogação e indexação são as seguintes: **título; subtítulo; sumário; resumo; introdução; prefácio; apresentação; títulos dos capítulos; bibliografia.**” Além da leitura de partes do documento, as bibliotecárias fazem uso de outras estratégias de leituras para auxiliá-las neste processo, a saber: consultas a outras bibliotecas (principalmente, a Biblioteca Nacional); sites especializados e bases de dados.

E considerando que a leitura documental é uma leitura técnica, que não exige que o profissional leia o documento integralmente, perguntou-se quanto tempo cada bibliotecária levaria ou leva para indexar um documento. As respostas revelam que o tempo varia muito de situação para situação, podem levar de 15 minutos a dias, devido a complexidade de cada documento e isso desencadeou outra pergunta: quais seriam os aspectos que mais dificultam o processo de leitura na indexação? (ver Tabela 1).

Tabela 1- Aspectos que dificultam a leitura documental

Aspectos	Dados por valor (1 a 5)/dificuldade
Falta de conhecimento da área	5; 2; 5; 5; 3
Barreira da língua	1; 1; 3; 1; 1

Assuntos muito técnicos	4; 4; 4; 1; 2
Tempo dedicado à leitura	4; 2; 2; 3; 5;
Execução de outras atividades	2; 5; 3; 3; 4

Fonte: Autoria própria.

Como a pergunta solicitava que as bibliotecárias enumerassem por grau de dificuldade (1 - mais difícil a 5 - menos difícil) os aspectos que interferiam no processo de leitura na indexação, obteve-se os seguintes resultados: a questão da barreira da língua, pela variação, é o aspecto que mais dificulta a leitura dos indexadores, principalmente, quando o documento está em inglês. Em seguida, está o tempo dedicado à leitura, devido a execução de outras atividades, que fica em terceiro lugar e, assuntos muito técnicos e falta de conhecimento da área ficam, respectivamente, em quarto e quinto lugar numa escala de 1 a 5 em grau de dificuldade.

A leitura é um processo complexo e envolve um leque de elementos que determinará sua eficácia. No mais, tais fatores serão responsáveis por diferenciar uma leitura da outra. Neste contexto Neves (2007, p. 2, grifo nosso), define a leitura de um texto como uma

[...] atividade cognitiva que requer esforço mental/cognitivo, envolvendo: percepção, memória, inferência e dedução. Além disto, é também um processo interativo, em que o leitor e o autor se relacionam mutuamente, sendo o texto o agente desta vinculação. O autor, quando produz o texto, estabelece uma coerência textual que é recuperada no momento da leitura. A reconstituição da coerência de um texto depende de processos cognitivos construídos pelo leitor/usuário para sua compreensão.

Em outras palavras, entende-se que a leitura é bem mais que a simples decodificação de símbolos, ou seja: é uma técnica ou método. A leitura é uma prática que atribui sentido ao texto, relacionando autor-texto-leitor, é uma troca mútua de significados. E para que o leitor compreenda com exatidão os sentidos do texto é necessário que ele associe tanto seus conhecimentos prévios que estão armazenados em sua memória com as informações do texto a ser lido. É a interação daquilo que já é compreensível e familiar ao leitor com a assimilação de uma informação nova.

Para a assimilação de uma nova informação é importante que o leitor desenvolva estratégias de leitura que vão desde a compreensão dos gêneros linguísticos, das estruturas textuais à organização lógica do texto proposta pelo autor. Neste contexto, entende-se que a leitura é uma atividade que envolve a interação autor, texto e leitor, onde "o leitor realiza um trabalho ativo de compreensão e interpretação do texto, a partir de seus objetivos, de seu conhecimento sobre o assunto, sobre o autor, de tudo o que sabe sobre a linguagem etc." (KOCH; ELIAS, 2012, p. 12). Assim, a lei-

tura é compreendida como uma atividade que vai além da decodificação de símbolos e signos linguísticos, e que envolve estratégias de seleção, antecipação, inferência e verificação, apresentando o leitor como um produtor ativo que atribui sentido ao texto. (KOCH; ELIAS, 2012).

Neste contexto, foi perguntado às bibliotecárias se concordavam que aspectos cognitivos, linguísticos e lógicos auxiliam no processo de leitura na indexação, todas responderam que sim, principalmente, no que tange ao processo de análise e interpretação dos documentos, o que facilitará na extração de termos que representam fielmente o documento. Cabe aqui, destacar um dos argumentos acerca desta questão, em que G/UEMA afirma que, “– [...] a indexação envolve aspectos diversos, não é uma atividade isolada, mas que relaciona-se com a ciência como um todo.” Com base no argumento de G/UEMA a indexação é uma área interdisciplinar, ou seja, tem buscado aporte teórico-metodológico em outras áreas, portanto, não é uma área isolada. Dentre estas áreas destaca-se a Psicologia Cognitiva que tem buscado estudar como o homem pensa, manipula, acumula e assimila a informação. Associado a este contexto, destacam-se os dispositivos assinalados pelas bibliotecárias, como sendo os que mais usam/acionam quando realizam a leitura documental, conforme Tabela 2:

Tabela 2- Dispositivos usados/acionados durante a leitura documental

Aspectos	Dados por valor (1 a 6)/relevância
Percepção	1; 1; 1; 1; 5
Memória	2; 2; 4; 2; 2
Representação mental do conhecimento	5; 5; 2; 6; 3
Conhecimentos prévios	4; 4; 3; 3; 1
Conhecimentos linguísticos	3; 3; 5; 4; 4
Experiências e expectativas de leitura	6; 6; 6; 5; 6

Fonte: Autoria própria.

Pela análise dos dados presentes na tabela 2, observa-se que o dispositivo que as bibliotecárias mais utilizam é a percepção. Considerando-se a afirmativa de Dias e Naves (2007, p. 26, grifo nosso) em que a leitura é um processo caracterizado por “[...] uma natureza interativa que depende tanto do texto quanto do leitor, consistindo em uma série de procedimentos coordenados que incluem operações perceptivas, linguísticas e cognitivas”, então, pode-se compreender que a percepção, associada ao conhecimento prévio e linguístico que o indexador tem acerca do assunto, lhe permitirá reconhecer, organizar e assimilar as informações con-

tidas no documento. Portanto, é de suma relevância a contribuição da percepção para a interpretação do documento.

Conforme Leiva (2012, p. 41-42, grifo nosso) a percepção é o canal pelo qual o indexador recebe a informação, pelo uso de

[...] três sentidos: visão, audição e tato (por exemplo: indexador cego que usa a leitura tátil). Pela visão, recebe o texto escrito (livro, artigo, lei etc.), a imagem fixa (fotografia, anúncio publicitário etc.) e a imagem em movimento (vídeo, filme), bem como objetos físicos sujeitos à indexação (escultura). Pela audição, recebe os sons articulados por sistemas linguísticos (discurso), sons por composição artística (música), sons pela interação social (ruídos urbanos, domésticos, festas etc.), sons mecânico industriais (cadeia de produção, perfurações etc.), sons da natureza (ruídos da floresta, ondas etc.) ou sons apelativos (relógios, timbres, sinos chamando para a oração etc.)

Mas, além da percepção, há ainda o uso da memória - que foi reconhecida como o segundo elemento mais usado neste processo. O indexador, através da percepção, percebe e identifica a informação do documento a ser indexada, essa nova informação é armazenada na memória (sensorial, de curto e longo prazo) e, dependendo da situação, poderá ser usada assim que o bibliotecário necessitar. Por exemplo: o bibliotecário indexador “pega” um livro para indexar sobre Literatura maranhense, é um livro novo que a biblioteca acabou de adquirir. Neste momento, a memória se apresenta de duas formas distintas: 1 - o bibliotecário vai se lembrar que apesar de ser um livro novo, a biblioteca já dispõe de outros exemplares que tratam de forma similar o assunto; 2 - suas lembranças acerca do tema, devido a leituras anteriores; de posse dessas informações o processo de análise ficará mais rápida e prática, bastando ao bibliotecário revisitar suas memórias anteriores. Parece um exemplo simples ou até mesmo banal, no entanto, percebe-se perfeitamente como a questão da memória está intrínseca neste processo e em tantos outros, pois, imagine se este bibliotecário não tivesse a mínima noção acerca do tema ou sequer lembrasse que já havia indexado o assunto antes, este seria um trabalho bem mais demorado.

Neste sentido, Leiva (2012, p. 42) salienta que é na memória que o indexador

Armazena dados relativos ao mesmo processo de indexação (tanto teóricos quanto práticos), adquiridos durante sua formação e através da experiência, dados da política de indexação da instituição aplicáveis sobre o uso de linguagem de indexação ou em relação aos usuários; ou dados da área que trabalha (Química, Direito, Meio ambiente etc.), entre outros.

Assim, graças à memória, o bibliotecário consegue codificar, armazenar e recuperar a informação guardada em sua mente que servirá para o processo de indexação.

Outros dispositivos assinalados pelas bibliotecárias são: os conhecimentos prévios; linguísticos e; as experiências e expectativas de leitura que ficaram, respectivamente, em terceiro, quarto e sexto lugar dentre os aspectos que mais as auxiliam no processo de indexação. No que tange aos conhecimentos prévios estes podem ser entendidos como um mecanismo pelo qual o leitor, a partir daquilo que adquiriu ao longo de sua vida, compreende o texto. Ele resulta da confluência de outros conhecimentos aprendidos pelo leitor e que são acessados durante o processo de leitura, a saber: o conhecimento linguístico, o conhecimento enciclopédico ou conhecimento de mundo e o conhecimento interacional, sem os quais a compreensão do texto ficaria comprometida.

De acordo com Kleiman (2013, p. 15), o conhecimento linguístico é “[...] aquele conhecimento implícito, não verbalizado, nem verbalizável na grande maioria das vezes, que faz com que falemos português como falantes nativos. É devido a este conhecimento que pode-se compreender a estrutura do texto, levando em consideração sua apresentação gramatical e lexical, em outras palavras, é possível “[...] compreender: a organização do material linguístico na superfície textual; o uso dos meios coesivos para efetuar a remissão ou sequenciação textual; a seleção lexical adequada ao tema ou aos modelos cognitivos ativados.” (KOCH; ELIAS, 2012, p. 40).

O conhecimento enciclopédico ou conhecimento de mundo “refere-se a conhecimentos gerais sobre o mundo - uma espécie de *thesaurus* mental - bem como a conhecimentos alusivos a vivências pessoais e eventos espaço-temporalmente situados, permitindo a produção de sentidos.” (KOCH; ELIAS, 2012, p. 42). Ou seja, para a compreensão do texto o leitor deve associar conhecimentos do uso da língua, bem como conhecimentos advindos de suas experiências e de seu convívio na sociedade que estão armazenados em sua memória de longo prazo, de forma a dar sentido ao texto. Trata-se de um “conhecimento estruturado (porque está ordenado), parcial (porque inclui apenas o que é mais genérico e previsível das situações) sobre um assunto, evento ou situação típicos.” (KLEIMAN, 2013, p. 26). Esse conhecimento é estruturado em forma de esquemas, que permite ao leitor ordenar mentalmente uma determinada situação ou informação, o que resulta na

[...] economia e seletividade na codificação de nossas experiências, isto é, no uso das palavras com as quais tentamos descrever para o outro as nossas experiências; podemos lexicalizar uma série de impressões,

eventos discretos através de categorias lexicais mais abrangentes e gerais e ficar relativamente certos que nosso interlocutor nos compreenderá. (KLEIMAN, 2013, p. 27).

Esses esquemas são as estruturas que o leitor cria em sua mente, de forma a ordenar e organizar todo o conhecimento adquirido ao longo de suas experiências e convívios com o meio social, o que lhe permite com mais eficiência compreender e dar sentido à leitura de um texto.

Neste sentido Neves (2006, p. 41), afirma que,

[...] a capacidade de compreensão textual está diretamente relacionada à capacidade do leitor de criar modelos mentais [esquemas] com base no significado declarado e não declarado pelo autor do texto. Desse modo, a construção de modelos mentais [esquemas] favorece a compreensão das palavras que lemos e suas combinações, possibilitando o entendimento do significado de um texto em dado contexto.

O conhecimento interacional, por sua vez, é “o conhecimento sobre as ações verbais, isto é, sobre as formas de inter-ação através da linguagem.” (KOCH, 2009, p. 48, grifo do autor). E engloba um leque variado de outros conhecimentos, dos quais destaca-se: o conhecimento ilocucional que refere-se à capacidade que o leitor tem em “[...] reconhecer os objetivos ou propósitos pretendidos pelo produtor do texto, em uma dada situação interacional.” (KOCH; ELIAS, 2012, p. 46). Está associado aos tipos de atos de fala, que podem ser verbalizados ou não, o que no segundo caso exigirá do leitor/interlocutor um grau de entendimento mais expressivo acerca dos objetivos do texto pretendido pelo autor/produtor de forma a assimilar o objetivo ilocucional da situação interacional a ele proposta (KOCH, 2009); o conhecimento comunicacional diz respeito “[...] à quantidade de informação necessária numa situação concreta para que o parceiro seja capaz de reconstruir o objetivo do produtor do texto; à seleção da variante linguística adequada a cada situação de interação e à adequação dos tipos de texto às situações comunicativas.” (KOCH, 2009, p. 49), em outras palavras, esse conhecimento envolve todos elementos necessários para o processo de comunicação humana; o conhecimento meta-comunicativo “é aquele que permite ao locutor assegurar a compreensão do texto e conseguir a aceitação pelo parceiro dos objetivos com que é produzido.” (KOCH; ELIAS, 2012, p. 52), em outro sentido, equivale aos mecanismos linguísticos utilizados pelo autor/produtor no texto de forma a evitar interferências e/ou ruídos na compreensão geral do texto pelo leitor; e o conhecimento superestrutural, por sua vez, “permite a identificação de textos como exemplares adequados aos diversos eventos da vida social.” (KOCH; ELIAS, 2012, p. 54). Ou seja, refere-se à diferenciação ou ordenação dos tipos textuais. Assim, é graças a este conhecimento que se sabe, por exemplo, quando um determinado texto refere-se ao gênero

jornalístico e não acadêmico, ou quando trata-se de uma poesia e não crônica ou fábula.

Compreende-se, assim, que o conhecimento linguístico, enciclopédico/de mundo e o interacional - que resultam no conhecimento prévio - são ativados sempre que se inicia o processo de leitura, mesmo que o leitor não perceba essa ativação, como forma de garantir a compreensão do texto. E, esse

conjunto de conhecimentos, socioculturalmente determinados e vivencialmente adquiridos, sobre como agir em situações particulares e realizar atividades específicas vêm a constituir o que chamamos de **'frames', 'modelos episódicos'** ou **'modelos de situação'**. (KOCH; ELIAS, 2012, p. 56, grifo dos autores).

Neste sentido, a leitura revela-se como uma atividade complexa que exige do leitor uma compreensão amplamente consolidada do contexto sociolinguístico, associando questões de cunho linguístico, como a estrutura e organização do texto, bem como incorporando a esta leitura traços de suas experiências e vivências cotidianas e de conhecimentos adquiridos e armazenados em sua memória ao longo de sua vida.

No que se refere ao aspecto, representação mental do conhecimento, as bibliotecárias o elegeram como sendo pouco utilizado neste processo, tal constatação pode ser explicada pela ausência de conhecimento, por parte das bibliotecárias, da influência que este tem sobre o processo de leitura. O uso da representação mental do conhecimento no processo de leitura justifica-se pelo seguinte fator, trata-se de um mecanismo pelo qual o indivíduo representa algo ou alguém em sua mente, o que lhe permite reconhecer ou identificar alguma coisa apenas por aquilo que tem representado em sua mente. Essas representações são responsáveis por desenrolar situações ou atividades cognitivas bastante complexas, como é o caso da leitura, haja vista que a leitura aciona, internamente, no indivíduo estruturas linguísticas, lógicas e cognitivas que associadas permitirão ao leitor realizar uma leitura sem grandes dificuldades, transformando-o em um leitor proficiente.

A ideia de estruturas cognitivas necessárias para a compreensão da leitura documentária, por exemplo, só é possível graças aos estudos advindos do campo da Psicologia Cognitiva, que ao longo dos anos tem se dedicado ao estudo de como o homem capta, assimila, codifica, armazena e recupera as informações necessárias para a solução de problemas, bem como para atividades de escrita, raciocínio, leitura, entre outras. Essas estruturas cognitivas são mecanismos que o sujeito desenvolve para representar mentalmente algo ou alguma coisa. Assim os conhecimentos prévios armazenados na memória de longo prazo garantem ao sujeito

cognoscente uma superioridade na compreensão de um texto, pois a leitura se torna mais fácil e rápida. Neste cenário, “a leitura passa a ser um processo cognitivo que depende da participação do leitor, que entra em cena com seu conhecimento prévio e participa da construção de significados; processo em que o texto passa a ser o mediador de comunicação entre o autor e leitor.” (BOSO et al., 2010, p. 29). Logo, o texto só ganha sentido a partir da compreensão do leitor, e essa compreensão dependerá de leitor para leitor, o que resultará em sentidos divergentes.

E considerando a leitura documental como parte integrante de um conjunto complexo e que dela dependerá a recuperação da informação, então considera-se que esta é uma etapa que envolve, dentre outros aspectos, a percepção, a memória e a representação mental do conhecimento para determinação do assunto contido no documento. Associado a este contexto, destaca-se a resposta de G/UEMA para a questão que se referia, justamente, a este ponto, a influência destes três fatores no processo de leitura. Segundo ela: “– Sim, a análise conceitual envolve sobretudo percepção e a representação mental do conhecimento, pois na leitura técnica do documento, a atenção, o conhecimento e o ponto de vista do indexador determinam a descrição temática do mesmo.”

Para expressar com mais propriedade a relação do processo de indexação com o processo cognitivo do bibliotecário (uso da percepção, memória e representação mental do conhecimento), utiliza-se o Quadro 2, abaixo representado:

Quadro 1 - Relação entre o processo cognitivo e a indexação

INDEXAÇÃO	PROCESSO COGNITIVO
Conhecimentos prévios do indexador: - Teoria e prática da indexação - Contexto da indexação a executar (necessidades dos usuários, políticas de indexação, condicionantes externos, etc.) - Assunto a ser indexado (História, Física, Direito...)	Memória de curto prazo
Percepção sensorial da informação: - Visual: texto, objeto físico, imagem - Auditiva: palavra, música, som ambiente	Memória sensorial: - Visual (memória icônica) - Auditiva (memória ecóica)

Análise do documento	Aplicação dos conhecimentos Prévios	Memória de curto prazo
	Reconhecimento de tipologias textuais	Memória de curto prazo Memória de longo prazo
	Identificação de Temas	Memória de curto prazo MLP Compreensão Estratégias redutoras de Informação
Conversão da linguagem natural para a linguagem controlada		Memória de curto prazo Memória de longo prazo
Armazenamento da indexação resultante na base de dados		Memória de curto prazo Memória de longo prazo

Fonte: Leiva (2012, p. 63)

Em atenção ao quadro 1, percebe-se que o processo de indexação está intimamente relacionado com os processos cognitivos, principalmente, no que se refere à memória (sensorial, de curto e longo prazo). É devido à memória sensorial, por exemplo, que o indexador consegue perceber a informação, seja pelo meio visual, onde apreende a informação do documento através do texto, do objeto físico e da imagem ou auditivo - pela palavra, música, som etc. Já na memória de curto prazo associa-se a atividades relacionadas ao conhecimento prévio, tais como: teoria e prática da indexação; contexto da indexação a executar e assunto a ser indexado. Neste contexto, salienta-se que os conhecimentos prévios armazenados na memória de curto prazo do bibliotecário são de suma relevância na representação do conteúdo de um documento, pois não será necessário que o indexador-leitor leia palavra por palavra ou linha por linha para compreender o teor do documento, basta apenas recorrer às informações estocadas em sua mente, o que lhe permitirá compreender com mais eficácia o conteúdo do documento a ser indexado, evitando que haja ruídos ou interferências indesejáveis durante este processo, o que refletirá diretamente na recuperação da informação pelo usuário. Quando se fala em conhecimentos prévios, refere-se àquilo que o bibliotecário conhece do documento a ser indexado, e isso compreende desde a tipologia linguística do texto até a forma lógica adotada pelo autor, para dar ao leitor uma visão do que será abordado.

Assim, entende-se que o processo de análise de assunto, onde se busca extrair os termos/descriptores que mais caracterizam e identificam o conteúdo do documento, deriva da interação entre as representações mentais que o indexador-leitor faz do mundo que o circunda e, que estão

armazenadas em sua memória sensorial, de curto e longo prazo, com o processo de assimilação e compreensão de uma nova informação advindos da percepção.

Para fechar a discussão acerca dos dispositivos usados ao longo do processo de leitura, foi proposto às bibliotecárias a seguinte afirmativa: “Durante a leitura de um texto são ativados esquemas variados, desde conhecimento de vocabulário, conhecimento da estrutura textual, do assunto, até conhecimento de mundo.” (FUJITA, 1999, p. 109) e perguntou-se se concordavam ou não e por quê. Todas concordaram com esta afirmativa, e dentre as respostas destacaram-se três, a saber:

– *Na leitura de um texto, analisa-se o conjunto, isto é, não apenas o assunto, mas todo o contexto que envolve aquele tema, época em que foi escrito (contexto social, histórico, político, etc.); além da análise linguística (interpretação textual, conhecimentos gramaticais, etc.) (G/UEMA);*

– *Sim, a leitura textual engloba a estrutura do texto, o conhecimento de vocabulário e até o conhecimento em geral que, para o bibliotecário, em sua prática de indexação vai caracterizá-lo como leitor profissional diferenciado do leitor comum. (B/UEMA);*

– *Penso, logo existo, indexo, logo recupero. [...]. (J/UFMA).*

Tais argumentos expressam o entendimento da análise de assunto como um processo complexo, que exige do bibliotecário uso de inúmeros aspectos, desde conhecimento de mundo, linguístico até cognitivos que, trabalhados conjuntamente, refletirão numa leitura eficaz e na extração adequada dos termos que qualificam o assunto do documento visando sua recuperação pelos usuários. Veja, por exemplo, o primeiro argumento, em que a bibliotecária afirma que o processo de leitura de um texto não envolve apenas o assunto em si, mas considera tantos outros fatores que permitirão ao bibliotecário realizar uma interpretação correta daquilo que o autor queria expressar no texto. E o uso destes fatores destinguirá o indexador-leitor dos demais leitores, conforme argumenta B/UEMA. Assim, uma boa leitura/interpretação do texto/documento refletirá na extração de termos significativos que resultarão na eficácia do processo de recuperação da informação, conforme argumenta J/UFMA.

Indexando textos: a prática do bibliotecário indexador

O segundo momento da pesquisa envolveu a realização de dinâmica cuja proposta era conhecer as estratégias de leitura usadas pelas bibliotecárias, as principais dificuldades apresentadas durante a leitura, bem como verificar a variação entre os termos extraídos dos textos.

a) Texto 1 - Fábulas: narrativa da literatura infantil: revivendo momentos de magia e encantamento

➤ Bibliotecária R da Biblioteca da Faculdade Santa Terezinha (CEST)

O primeiro contato da bibliotecária R/CEST foi com o texto 1, onde explicou que para extrair os descritores dos textos, ela considerava alguns aspectos, dentre eles: o reconhecimento dos autores, o título do artigo, o nome da revista, o resumo e a introdução. Conforme R/CEST, o resumo é uma ótima fonte para se extrair os descritores, pois, em sua maioria são bem fundamentados e completos. No que se refere a este primeiro texto R/CEST teve facilidade em extrair os termos, principalmente, porque era um tema familiar a ela devido já ter experiências anteriores, pois havia trabalhado por muito tempo com literatura infantil em uma biblioteca.

Ao extrair os termos R/CEST confidenciou que tenta esgotar todas as possibilidades de descritores que o texto possa apresentar, pois, segundo ela, o objetivo principal é atender às necessidades dos usuários e evitar que haja ruídos entre o que está sendo indexado e recuperado pelo usuário. Neste contexto, os termos extraídos foram: **Fábulas; Literatura infantil e Narrativa - Literatura infantil.**

➤ Bibliotecária I da Biblioteca Central da Universidade Federal do Maranhão (UFMA)

A primeira estratégia usada por I/UFMA foi verificar se o texto possuía palavras-chave, visto que é a principal fonte utilizada por ela para extrair os termos, no entanto, o texto 1 não apresentava palavras-chave, portanto, I/UFMA precisou realizar uma leitura mais demorada e dinâmica do texto. Assim, ela folheou o texto, leu as partes que mais lhe interessava, considerou a conclusão e por fim selecionou os seguintes descritores: **Literatura infantil; Estudo de narrativas** (estes dois termos foram extraídos considerando apenas o resumo) e; **histórias infantis** (este termo foi extraído após leitura do texto). No mais, I/UFMA revelou que faz uso de um vocabulário controlado da própria instituição, pois busca sempre elencar os termos mais específicos do texto e manter certa uniformidade com documentos já indexados que tratam da mesma temática.

- Bibliotecária B da Biblioteca da Universidade Estadual do Maranhão (UEMA)

A bibliotecária B/UEMA realizou a análise do texto 1 considerando apenas a leitura do título e do resumo, de onde extraiu o assunto principal - **Literatura infantil e fábulas**. B/UEMA sempre busca extrair os termos conforme a necessidade de seus usuários e da própria biblioteca e, para manter certa uniformidade e especificidade, ela sempre busca comparar os termos extraídos com os termos presentes na Classificação Decimal Universal (CDU).

- Bibliotecária J da Biblioteca Central da Universidade Federal do Maranhão (UFMA)

A bibliotecária J/UFMA teve bastante facilidade em extrair os termos do texto 1, pois considerou apenas o título e o resumo. Foi mais prática, não chegou a folhear todo o texto, uma vez que considerou o tema fácil. Os descritores extraídos foram: **literatura infanto-juvenil** e **fábulas**.

b) Texto 2 - Humanização do atendimento em saúde: conhecimento veiculado na literatura brasileira de enfermagem

- Bibliotecária R da Biblioteca da Faculdade Santa Terezinha (CEST)

No segundo texto, as estratégias de leitura foram as mesmas utilizadas na análise do texto 1: leitura do título, do resumo, da introdução, reconhecimento dos autores, da revista etc., no entanto, R/CEST demorou um pouco mais para realizar a extração dos termos, haja vista que o tema era pouco conhecido por ela. Assim, foi preciso fazer uma leitura mais criteriosa a fim de ela extrair os termos mais significativos. Durante a leitura, foi possível observar que, R/CEST buscou identificar a proposta principal do artigo, bem como reconhecer a estrutura lógica deste: introdução, metodologia, assunto principal, conclusão etc., a fim de se familiarizar com a temática, já que se tratava de um assunto pouco conhecido por ela. Após esta leitura mais detalhada, R/CEST conseguiu extrair os seguintes descritores: **Humanização**, que segundo ela representa o assunto principal do artigo; **Atendimento em saúde** e **Produção científica** (assunto principal) - **enfermagem** (assunto secundário), neste último descritor, R/CEST explicou que busca sempre destacar o assunto principal e caso este assunto precise de um assunto complementar, ela utiliza um segundo termo, o qual chama de assunto secundário e, esses termos são ligados por um hífen.

Em geral, R/CEST faz uma leitura mais rápida/dinâmica/técnica e superficial de assuntos recorrentes ao seu dia a dia ou de seu conhecimento prévio. E no caso de temas mais direcionados e específicos, ela faz uma

leitura mais demorada e minuciosa a fim de se familiarizar com o texto e assim extrair os termos mais adequados.

- Bibliotecária I da Biblioteca Central da Universidade Federal do Maranhão (UFMA)

I/UFMA reservou um pouco mais de tempo para a leitura deste texto 2, mesmo identificando a presença das palavras-chave, pois I/UFMA sempre lê o texto como forma de verificar se os termos utilizados pelos autores realmente traduzem a conteúdo do artigo, assim I/UFMA constatou que um termo estava presente tanto no título quanto ao longo do texto, no entanto, não havia sido indexado pelos autores. E segundo I/UFMA este representava o assunto principal do texto, assim ela considerou todas as palavras-chave usadas pelas autores (com exceção do termo - prestação de cuidados de saúde) e acrescentou humanização em enfermagem. Assim os descritores extraídos foram: **humanização em enfermagem; serviços de saúde e enfermagem.**

Em comparação com R/CEST, I/UFMA foi bem mais técnica, utilizou pouco a questão do conhecimento prévio sobre o assunto, focando mais na estrutura do texto, realizando uma análise de assunto mais isolada e tecnicista. No mais, foi possível observar certa dificuldade de I/UFMA em definir os termos, mesmo considerando que o conhecimento sobre a temática lhe auxilia nesta atividade.

- Bibliotecária B da Biblioteca da Universidade Estadual do Maranhão (UEMA)

Nesse segundo texto, B/UEMA seguiu as mesmas estratégias usadas no primeiro texto, ou seja, leitura do título, do resumo etc. Durante a leitura confidenciou que quando não conhece o assunto ela sempre recorre a outras bibliotecas para verificar se o título em questão já foi indexado ou então ela busca auxílio em *sites*, para ler resenhas, por exemplo. Igualmente a R/CEST, B/UEMA também busca esgotar todas as possibilidades de descritores, pois segundo ela quanto mais termos forem extraídos mais rápida e prática tornar-se-á a recuperação desta informação. Os termos extraídos foram: **humanização; enfermagem; produção científica e; serviço de saúde.**

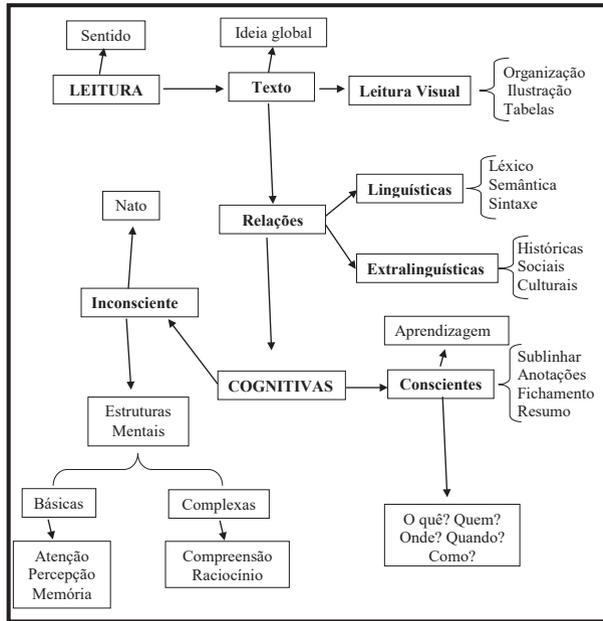
- Bibliotecária J da Biblioteca Central da Universidade Federal do Maranhão (UFMA)

No texto 2, foi possível observar que J/UFMA não conhecia a temática, por isso, precisou folhear todo o artigo, considerando a leitura do título, do resumo, da introdução e dos tópicos do artigo, a fim de se familiari-

zar com o assunto. Só assim conseguiu extrair os seguintes termos: saúde pública - atendimento; humanização no atendimento e espaço físico.

Conforme mostra os relatos acima, as principais partes do documento utilizadas pelas bibliotecárias para extrair os termos foram: título, resumo, introdução, tópicos do texto e conclusão, de acordo com o que sugere a literatura. Mas, além dessas estratégias, foi possível constatar, empiricamente falando, a questão das estruturas mentais ou do esforço cognitivo (NEVES, 2007), que as bibliotecárias desenvolveram para realizar a atividade de leitura, visto que a leitura é um sistema complexo que articula conhecimentos linguísticos e cognitivos, conforme ilustra Figura 5.

Figura 2 – Relação linguística e cognitiva do processo de análise conceitual



Fonte: Adaptado de Pereira (2013). Aula do curso de Biblioteconomia da UFMA/Disciplina: Análise Temática da Informação.

Assim, ao realizar a leitura do texto, o bibliotecário deve lhe atribuir sentido e ser capaz de extrair sua ideia principal/global. Neste sentido, supõe-se que a leitura “deve ser balizada por regras, padrões, esquemas, enfim, pela utilização de estratégias de leitura adequadas. [Onde o texto] é colocado como autônomo e o levantamento dos seus possíveis sentidos é reduzido a mecanismos, estratégias de leitura.” (LUCAS, 2000, p. 78). Dentre estas estratégias destacam-se a forma como o texto está or-

ganizado, portanto, sugere-se a leitura do título, do resumo, da introdução, dos tópicos, da conclusão, de ilustrações ou tabelas (leitura visual) etc. No mais, compreende-se que o bibliotecário deve buscar associar ou correlacionar aspectos linguísticos (léxico, semântico e sintaxe) e extralinguísticos (aspectos históricos, sociais e culturais - conhecimentos prévio). Mas, o principal diferencial no processo de leitura é o reconhecimento e uso, por parte dos bibliotecários, das estruturas cognitivas que podem ser conscientes - que associa-se aos mecanismos utilizados pelo indexador, como questões do tipo: o quê? quem? onde? quando? e como?, o que lhe permitirá identificar a ideia central do documento e dele extrair os termos que melhor representem o documento original. Ou inconscientes - uso da percepção e da memória, que auxiliam na compreensão no conteúdo tratado no documento.

As principais dificuldades apresentadas pelas bibliotecárias durante o processo de análise conceitual refere-se à falta de conhecimento do assunto. No que tange, à variação de termos extraídos, observou-se que segue certa padronização, mesmo tratando-se de indexadores diferentes, conforme Quadro 2:

Quadro 2 - Principais descritores extraídos pelas bibliotecárias

Texto 1 - Fábulas: narrativa da literatura infantil: revivendo momentos de magia e encantamento	
Bibliotecárias/ Instituição	Descritores
R/CEST	Fábulas; Literatura infantil; Narrativa-Literatura infantil.
B/UEMA	Fábulas; Literatura infantil;
I/UFMA	Literatura infantil; Estudo de narrativa; Histórias infantis.
J/UFMA	Fábulas; Literatura infanto-juvenil.
Texto 2 - Humanização do atendimento em saúde: conhecimento veiculado na literatura brasileira de enfermagem	
R/CEST	Humanização; Atendimento em saúde; Produção científica-enfermagem.
B/UEMA	Humanização; Enfermagem; Produção científica; Serviço de saúde
I/UFMA	Humanização em enfermagem; Serviços de saúde; Enfermagem.
J/UFMA	Saúde pública - atendimento; Humanização no atendimento; Espaço físico.

Fonte: Autoria própria com dados obtidos mediante dinâmica - Indexando textos.

Conforme quadro 2, observa-se que no caso do texto 1 os descritores extraídos são praticamente os mesmos, com exceção da bibliotecária I/UFMA que não utilizou a palavra fábulas, mas histórias infantis. E no tex-

to 2, o que se percebe é a variação do termo *humanização*, que representa o assunto principal do texto, e que foi citado só e ligado a outras palavras, mas foi o termo mais utilizado pelas bibliotecárias.

O processo de indexação é uma atividade subjetiva que depende da ação dos bibliotecários, mas em muitos casos esta ação está sujeita a interferência de fatores intelectuais – que envolvem o conhecimento prévio, aspectos lógicos, linguísticos e cognitivos deste profissional com relação ao assunto a ser indexado, e físicos. No mais, compreende-se que a indexação deve ser feita respeitando a necessidade de cada biblioteca e de seus respectivos usuários. Assim, compreende-se que “indexador precisa conhecer os interesses destes usuários e, na escolha dos termos deve-se considerar os vários aspectos. Por isso, a mesma fonte de informação pode estar indexada de diferentes modos, focando interesses diversificados dos usuários.” (VARELA; BARBOSA, 2007, p. 123). Ou, como sugere Lucas (2000, p. 76), “[...] o bibliotecário sobressai como um leitor guiado por uma metodologia de trabalho que o sujeita às coerções institucionais, para que sua leitura seja a do interesse dos usuários da instituição à qual sua leitura serve.” Tal afirmativa, corrobora com o pensamento de algumas das bibliotecárias pesquisadas, que consideram sempre as necessidades de seus usuários e da sua instituição ao realizar a indexação de documentos, pois buscam tornar visível e útil todo e qualquer documento informacional.

ASPECTOS CONCLUSIVOS: entre o fazer e a teoria

A indexação, ou melhor, a análise de conteúdo é uma atividade eminentemente humana e, portanto, subjetiva, que depende da ação do bibliotecário indexador. Dentre os fatores envolvidos nesta atividade destacam-se os conhecimentos prévios, as estratégias de leitura, os conhecimentos linguísticos, semióticos, lógicos e principalmente, cognitivos.

Diante das discussões tratadas ao longo deste capítulo, compreende-se que a indexação é um processo complexo, principalmente no que tange à leitura documental realizada na primeira etapa desse processo – a análise conceitual. Esta é uma atividade aparentemente simples, mas que após uma análise detalhada percebeu-se que trata-se de algo bem mais que uma simples técnica, é a confluência de várias áreas, a saber: a linguística que cruza-se com esta atividade, a partir do momento em que se percebe que ambas trabalham com a linguagem, e que os termos extraídos, mediante a leitura técnica ou documental, como preferir, são unidades ou signos linguísticos que carregam consigo o sentido e o significado do documento (aspectos semióticos), só que de maneira codificada para propiciar a representação, organização e recuperação da infor-

mação pelos usuários. No mais, compreende-se que graças aos conceitos advindos da Linguística, o bibliotecário indexador consegue reconhecer as estruturas textuais do documento, bem como as relações sintáticas e semânticas dos termos ou descritores que melhor representam o assunto abordado ao longo no documento. A lógica, também oferece à indexação aportes teóricos que residem nos raciocínios dedutivo e indutivo, o que permitem aos bibliotecários decidir a melhor estratégia de leitura documental. Estas áreas, Linguística, Semiótica e Lógica, juntamente com a Psicologia Cognitiva caracterizam o processo de indexação enquanto área interdisciplinar.

A Psicologia Cognitiva vem ao longo dos anos se dedicando aos estudos que cercam a estrutura mental do indivíduo, de forma a entender como este percebe, apreende, compreende, assimila, armazena e recupera a informação. Assim, destacam-se estudos voltados para a percepção, a memória e a representação mental do conhecimento. Pautado nestes três pilares é que buscou-se fundamentar e caracterizar o objetivo deste estudo, que foi conhecer e analisar os aspectos cognitivos que auxiliam o bibliotecário indexador no processo de leitura realizado na indexação.

A priori, fora questionado: “Até que ponto os aspectos linguísticos, lógicos, semióticos e cognitivos podem auxiliar este profissional no processo de representação, organização e recuperação da informação pelos usuários?” E como resposta, constatou-se que tais fatores estão presentes do início ao fim do processo de indexação. Tal constatação baseou-se em dados obtidos mediante aplicação de questionários e de uma dinâmica confrontando com a literatura.

Observou-se que a leitura documental é uma atividade complexa que vai além da decodificação de símbolos e sinais gráficos e visuais presentes nos textos e documentos. Ela pressupõe a ativação de aspectos cognitivos que envolve aquilo que o bibliotecário indexador percebe (sistema sensorial), assimila, compreende e armazena em sua memória (sensorial, de curto e longo prazo), a partir de representações mentais de tudo que o rodeia.

Identificou-se que as bibliotecárias ainda apresentam dificuldades com relação a análise de assunto, principalmente quando o documento encontra-se em outra língua – o inglês, por exemplo, e com relação a documentos que tratam de assunto muito técnico; além disso o tempo destinado à leitura é insuficiente, visto o acúmulo de outras atividades pelos indexadores que é consequência da insuficiência ou mesmo ausência de recursos humanos para desenvolver tais atividades. Em geral, o que se constata na realidade é que há um ou dois indexadores por instituição

universitária no Maranhão, com exceção da Biblioteca Central da Universidade Federal do Maranhão em que há um quantitativo bem expressivo de indexadores e esse quantitativo pode ser rapidamente justificado, uma vez que todo processamento técnico é feito por estas bibliotecárias, ou seja, todos os documentos que abastecem os diversos polos da UFMA são tratados pelos indexadores da Biblioteca Central, o que de certa forma indica o acúmulo de atividade e insuficiência de pessoal que reflete diretamente na prática de indexação. No caso da Biblioteca da UEMA, constatou-se a presença de duas bibliotecárias para a prática de indexação: uma é responsável pela indexação de periódicos e a outra de livros. No CEST a situação se repete, há apenas duas bibliotecárias que se revezam entre processos técnicos e atendimento ao usuário.

Neste contexto, o que se percebe é a falta de relevância dada aos processos técnicos e, principalmente, à indexação. Assim, muitos bibliotecários acabam dando mais importância para outras atividades, devido a ausência de pessoal como já salientado, e esquecem da importância que a indexação tem para a recuperação da informação, pois entende-se que uma informação só será recuperada desde que seja devidamente tratada. Neste sentido, compreende-se que para a leitura documentária ser eficiente, de qualidade e que garanta a eficácia na identificação, seleção de termos, tradução e, conseqüentemente, na recuperação da informação é importante considerar alguns pontos cruciais e imprescindíveis para a consolidação deste processo, como: melhores condições de leitura, com: ambiente agradável e organizado sem acúmulo de documentos, iluminação e ventilação adequada, mais tempo dedicado à leitura, com a contratação de mais bibliotecários, entre outros. Com relação às dificuldades apresentadas pelas bibliotecárias sugere-se: capacitação e atualização contínua, aprimoramento com relação às línguas estrangeiras e a utilização de normas e procedimentos que visam a eficácia do processo de leitura documental, dentre eles: a Norma 12.676, da Associação Brasileira de Normas Técnicas de 1992 etc.

Para minimizar as dificuldades referentes a assuntos muito técnicos é necessário que as bibliotecárias mantenham contato com especialistas da sua área de atuação bem como de outras áreas do conhecimento. Sugere-se, ainda, a formalização e consolidação de uma política de indexação, visto que nas três instituições pesquisadas não existe uma política, o que há é a repetição de práticas cotidianas que acabam passando por política. No mais, compreende-se que uma política documental garante a uniformidade e padronização dos termos extraídos dos documentos e minimiza ruídos ou divergências neste processo.

Diante deste contexto, salienta-se que o bibliotecário deve se apropriar, não só de conhecimentos exigidos pela área da Biblioteconomia, mas deve se valer de muitos outros conhecimentos, que são imprescindíveis para a sua prática profissional, dentre eles destacam-se: o conhecimento linguístico e semiótico que se refere aos meios pelos quais o leitor compreende o texto, sua estrutura e identifica sua essência, mesmo que esta não seja devidamente declarada pelo autor; o conhecimento lógico que auxilia o leitor na escolha da melhor estratégia de leitura e os conhecimentos acerca dos esquemas mentais que o indivíduo cria quando realiza a leitura de um texto, por exemplo, que advém da área da Psicologia Cognitiva. E por entender que aspectos cognitivos é algo nato do indivíduo, ou seja, nasce com ele, o que lhe falta é reconhecer sua utilidade para o desenvolvimento de atividades no campo da Representação e Organização da Informação, bem como no desenrolar de muitas outras atividades, pois entende-se que conceitos como: percepção, memória e representação mental do conhecimento fazem parte da construção social e humana de todo e qualquer indivíduo.

Por fim, constatou-se que o bibliotecário é um sujeito cognoscente envolvido em um sistema cognitivo bem mais complexo, que por vezes passa totalmente despercebido, mas que interfere e auxilia o indexador na prática de análise conceitual (leitura documentária).

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12676**: métodos para análise de documentos - determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro: ABNT, 1992.

BOCCATO, Vera Regina Cassari. O contexto sociocognitivo do indexador no processo de representação temática da informação. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, v. 17, n. esp. 1, p. 67-86, 2012.

BOSO, Augiza Karla; GARCIA, Daniela; RODRIGUES, Michele de Britto; MARCONDES, Pollyne. Aspectos cognitivos da leitura: conhecimento prévio e teoria dos esquemas, **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis, v. 15, n. 2, p. 24-39, jul./dez., 2010.

CUNHA, Anna Maria Marques. **Análise documentária**: a análise da síntese. Brasília: IBICT, 1987.

CUNHA, Miriam Vieira da. O profissional da informação e o sistema das profissões: um olhar sobre competências. **Ponto de Acesso**, Salvador, v. 3, n. 2, p. 94-108, ago. 2009.

DIAS, Eduardo Wense; NAVES, Madalena Martins Lopes. **Análise de assunto**: teoria e prática. Brasília: Tesouro, 2007. (Estudos Avançados em Ciência da Informação, 3).

FUJITA, M. S, L.: RUBI, M. P. Um modelo de leitura documentária para indexação de artigos científicos: princípios de elaboração e uso para a formação de indexadores. **DataGramaZero**, Rio de Janeiro, v. 7, n. 3, p. 1-19, jun. 2006.

FUJITA, Mariângela Spotti Lopes. A leitura do indexador: estudo de observação. **Perspectivas em Ciência da Informação**, v. 4, n. 1, 1999.

KLEIMAN, Angela. **Oficina de leitura** - teoria e prática. 15. ed. Campinas, SP: Pontes Editores, 2013.

KOCH, I. G.V. **A coesão textual**. 21.ed. São Paulo: Contexto, 2009.

KOCH, I. V. **Ler e compreender**: os sentidos do texto. 3. ed. São Paulo: Contexto, 2010.

LEIVA, Isidoro Gil. Aspectos conceituais da indexação. In: LEIVA, Isidoro Gil; FUJITA, Mariângela Spotti Lopes [editores]. **Política de Indexação**. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012.

LUCAS, Clarinda Rodrigues. **Leitura e interpretação biblioteconomia**. Campinas: Editora da UNICAMP, 2000.

NEVES, D. A. de B. Representação Temática da Informação e mapas cognitivos: interações possíveis. **Informação & Sociedade: Estudos**, João Pessoa, v. 22, p. 39-47, Número Especial 2012.

NEVES, Dulce Amélia de Brito, Leitura e metacognição: uma experiência em sala de aula, **Enc, Bibli: R. Eletr, Bibliotecon. Ci. Inf.**, Florianópolis, n. 24, p. 1-9, 2007.

NEVES, Dulce Amélia. Ciência da informação e cognição humana: und abordagem do processamento da informação. **Ci. Inf.**, v. 35, n. 1, jan/abr. 2006.

NOVELLINO, Maria Salet Ferreira. Instrumentos e metodologias de representação da informação. **Informação&Informação**, V. 1, n. 2, 1996.

OLIVEIRA, D. S. de. **Estratégias para o ensino de indexação**: o uso do protocolo verbal interativo como recurso pedagógico de indexadores aprendizes. 2005. 225 f. Monografia (Graduação em Biblioteconomia) -

Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2005.

RUBI, Milena Polsinelli. Os princípios da política de indexação na análise de assunto para catalogação: especificidade, exaustividade, revocação e precisão na perspectiva dos catalogadores e usuários. In: FUJITA, Mariângela Spotti Lopes (Org.). **A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias**. São Paulo: Cultura Acadêmica, 2009.

VARELA, Aida; BARBOSA, Marilene Lobo Abreu. Aplicação de teorias cognitivas no tratamento da informação. **Revista Brasileira de Biblioteconomia e Documentação**, v. 3, p. 116-128, 2007.



O PROCESSO DE CLASSIFICAÇÃO BIBLIOGRÁFICA NA ÁREA JURÍDICA: análise comparativa da CDU e CDDir nos domínios do Direito Administrativo e Constitucional

Cintia da Luz Santos

1 INTRODUÇÃO

As bibliotecas são instituições responsáveis pela guarda, organização, localização, e disseminação da informação, atividades que se constituem em uma árdua tarefa, em virtude da grande massa documental. A produção bibliográfica cresceu espantosamente nos últimos anos em todos os campos do conhecimento, principalmente na jurídica que é uma das áreas que está em constante atualização. Assim, nas bibliotecas voltadas para esse campo, toda essa produtividade, precisa ser devidamente selecionada, para saber quais os documentos relevantes, dar o tratamento adequado, e disponibilizá-los de maneira eficaz e satisfatória para o usuário. Para isso é necessário que haja um profissional capacitado, ou seja, um bibliotecário, que acompanhe as inovações tecnológicas, as ideias da sociedade ou comunidade na qual atua.

Uma das atividades desenvolvidas pela biblioteca, que proporciona organizar e recuperar a informação, com vista a torná-la acessível para os mais diversos fins é a classificação bibliográfica. A função que a classificação desenvolve na organização do conhecimento, é, de fato, uma das mais importantes, indispensáveis para que bibliotecas alcancem seu objetivo: satisfazer as necessidades de informação de seus usuários (LANGRIDGE, 2006).

A classificação é a maneira de agrupar características semelhantes de objetos ou seres. A classificação bibliográfica, na Biblioteconomia, serve para classificar os documentos de acordo com a área do conhecimento e, para isso, são utilizados Sistemas de Classificação, dando-se destaque

para a Classificação Decimal de Dewey (CDD) e Classificação Decimal Universal (CDU), (PIEDADE, 1983). Na CDD o domínio do Direito está classificado sob o número 340 dentro da classe de Ciências Sociais; na CDU recebe a numeração 34 que também pertence à classe das Ciências Sociais.

A área jurídica tem uma classificação própria para distinguir seus vários ramos (administrativo, constitucional, civil, tributário, processual), que é a Classificação de Direito (CDDir), criada em 1948, com o objetivo de proporcionar um sistema de classificação adaptado à realidade do Direito brasileiro

Apesar de a área Jurídica ter uma classificação própria, a essência do tratamento da informação vem de métodos tradicionais já estabelecidos na área da Biblioteconomia. A classificação bibliográfica que, de um modo geral, é uma atividade desenvolvida pelo profissional bibliotecário, oferece método para o processo de tratamento da informação, que busca o aprimoramento de suas técnicas na tentativa de melhorar a representação documentária para descrição e recuperação de recursos informacionais.

Isto posto, questiona-se, no entanto, o seguinte:

- a) Como estão representados os subdomínios do Direito Administrativo e Constitucional na CDU e CDDir?
- b) Qual o impacto da classificação no processo de busca e recuperação da informação na área jurídica, especificamente no subdomínio do Direito Administrativo e Constitucional?
- c) Quais os distanciamentos e as aproximações entre as representações na classificação dos subdomínios do Direito Administrativo e Constitucional na CDU e CDDir?

Assim o objetivo geral deste estudo é descrever e analisar a representação temática do domínio jurídico especificamente dos subdomínios do Direito Administrativo e Constitucional na CDU e CDDir, adotados pela Biblioteca do TJ-MA e pela Biblioteca Central da UFMA, na perspectiva de compreender a atividade da Classificação Bibliográfica na área jurídica. Para atingir esse objetivo maior foram traçados como objetivos específicos:

- a) Caracterizar e analisar os diferentes tipos de Classificação Bibliográfica;
- b) Conhecer as classes 34 da CDU e 340 da CDDir;

- b) Identificar as semelhanças e divergências nas CDU e CDDir na organização e estruturação dos conceitos nos subdomínios do Direito Administrativos e Constitucionais.

As estratégias adotadas nesta pesquisa, para alcançar seus objetivos, são essenciais, pois como enfatiza Braga (2007, p. 18) “Para a ciência não estão em jogo apenas os resultados da pesquisa, mas quais foram os meios e procedimentos adotados pelo pesquisador [...]”. A metodologia consiste na descrição detalhada, que o pesquisador faz para desenvolver sua pesquisa.

Nessa direção foram adotados para a realização do estudo, a pesquisa exploratória de natureza descritiva que, segundo Prestes (2003, p. 26), “[...] tem como objetivo proporcionar maiores informações sobre o assunto que vai ser investigado[...]”. Essa etapa acontece antes do desenvolvimento do trabalho, para aprimorar as ideias e proporcionar uma maior familiaridade com a questão norteadora da investigação

Em se tratando da pesquisa descritiva Gil (2007, p. 42) esclarece que “[...] tem como objeto primordial a descrição das características de determinada população ou fenômeno ou, então, o estabelecimento de relações entre variáveis.” Ou seja, descrever a prática, a relação ou conexão entre os fatos. Neste estudo foram observadas as variações entre as classificações CDU e CDDir, e suas relações com os termos selecionados, a saber: Licitação e Contratos, Regime Jurídico do Servidor Público, Direitos e garantias fundamentais, Poderes do Estado e Atos administrativos.

Na coleta de materiais foi utilizada a pesquisa bibliográfica, que “[...] explica um problema a partir de referências teóricas publicadas em artigos, livros, dissertações e teses.” (CERVO; BEVIAN; SILVA; 2007, p.60). Ela dá credibilidade para a pesquisa, devido à base teórica. Assim o estudo se deu primeiramente com a leitura e fichamento de livros, dissertações, artigo e tese, entre outros que abordam a temática da pesquisa. A revisão de literatura realizada nesta pesquisa, está pautada em autores como Piedade (1983) estudiosa da Classificação Bibliográfica, Anjos (2008), Carvalho (2002) que desenvolveu a CDDir (Classificação Decimal de Direito), Guinchat e Menou (1994), Langridge (2006), Dicionário e Enciclopédia entre outros, que tratam sobre a temática ora estudada.

A análise se deu por meio da estruturação dos conceitos da área jurídica no Sistema de Classificação Decimal Universal (CDU) e no Sistema de Classificação Decimal de Direito (CDDir) a partir do acesso aos sistemas de Automação adotados pela Bibliotecas Central da UFMA e pela Biblioteca do TJ.

A Biblioteca Central tem como missão apoiar a Universidade Federal do Maranhão nas funções de ensino, pesquisa e extensão, bem como preservar a informação, possibilitando a sua recuperação e difusão, através de serviços e produtos ofertados à comunidade na qual está inserida, absorvendo para si a responsabilidade de proporcionar a disseminação de conhecimentos capazes de provocar mudanças individuais, sociais, políticas e econômicas – (UFMA, 2013) e onde os usuários são de diferentes áreas do conhecimento. Na UFMA é utilizada a CDU, nesta instituição foi consultado o Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA).

A Biblioteca do Tribunal de Justiça (TJ) tem a missão de “Disponibilizar ao usuário o acesso às fontes de informação jurídica através da aquisição, organização e disseminação de conhecimentos, contribuindo para a eficiência na prestação jurisdicional.” (PODER JUDICIÁRIO, 2013), onde os usuários são específicos da área jurídica. Utiliza a CDDir e o sistema adotado é o Sistema de Automação de Bibliotecas (SIABI).

Dentre outros aspectos que serão abordados ao longo desta monografia, dá-se ênfase à Classificação do conhecimento, como fundamento basilar para compreender o problema de pesquisa, acima enunciado. Além de proporcionar o cumprimento do objetivo geral desta pesquisa, tal abordagem fornece os elementos que constituem o corpo deste estudo. a “Classificação Bibliográfica na Área jurídica.”

No passeio feito pela classificação do conhecimento, é dado, nesta seção, destaque para as bases da Classificação Filosófica, isto é, para o trabalho dos filósofos gregos que a usaram como forma de organizar o quadro de conhecimento da época e, por conseguinte, conhecer a origem do cosmo e do universo. A Classificação bibliográfica e seu uso em bibliotecas também são discutidos, em busca da identificação de suas razões, suas implicações e seus desdobramentos possíveis.

Após assumir a perspectiva da Classificação Bibliográfica como atividade essencial no processo de organização e recuperação das informações pela biblioteca, são estabelecidos os distanciamentos e as aproximações entre a Classificação Decimal Universal (CDU) e a Classificação Decimal de Direito (CDDir)

Por se tratar de uma área extremamente importante para a sociedade, pois é a área voltada para a conduta do cidadão, a área jurídica está relacionada diretamente ao contexto social e sua produção, representada em fontes como, doutrina, jurisprudência e legislação, torna-se necessário conhecer o contexto jurídico e a hierarquização dos conceitos nos sistemas de Classificação Bibliográfica.

Um estudo sobre esse assunto se justifica por proporcionar o conhecimento e compreensão desse importante ramo do saber e sua representatividade, refletida nos instrumentos de classificação adotados no fazer biblioteconômico, no processo ininterrupto de construção e reconstrução do conhecimento humano. Ademais, o interesse pelo tema surgiu pela inquietação pessoal e acadêmica em relação à representação da área.

Espera-se que o estudo possa desencadear a investigação tanto horizontal como vertical da temática, no sentido de contribuir para o aprofundamento da questão da organização e recuperação da informação.

UM PASSEIO PELA CLASSIFICAÇÃO DO CONHECIMENTO: da gênese ao uso

Desde a Antiguidade, a necessidade de organizar os objetos e conhecimentos, fez com que o homem começasse a separar e agrupar as coisas de acordo com suas semelhanças, O ato de classificar é inerente ao ser humano, que classifica de forma consistente ou não a partir de variáveis como cor, tamanho entre outros, tudo que está em sua volta, seja nos ambiente de casa, trabalho ou lazer. Ela é necessária pois “Sem Classificação, efetivamente não poderia haver nenhum pensamento humano, ação e organização que conhecemos.” (LANGRIDGE, 2006, p.11).

A palavra classificação, segundo Piedade (1983, p.17) “[...] vem do latim do *classis*, que designava os grupos em que se dividia o povo romano.” De modo geral, o termo Classificar pode ser usado para indicar o processo de organizar objetos, abstrações e/ou ideia de acordo com suas características. Souza (2010, p.13) o conceitua como “[...] processo de reunir objetos, ideias ou seres, em grupos de acordo com seu grau de semelhanças.”

A humanidade divide tudo em grupos ou classes, para compreender o que está em sua volta, Piedade (1983, p.19) concebe classe como sendo “[...] um conjunto de coisas ou ideias que possuem um ou vários atributos, predicados ou qualidades em comum.”

As classes são definidas através das características ou ideias, como alto ou baixo, amarelo ou azul; se é natural ou se é artificial entre outras. Para Anjos (2008, p. 235) característica é “[...] uma propriedade distintiva, pela qual uma classe é definida em um esquema de classificação”. As categorias são as classes maiores, “[...] as grandes classes, os grandes tipos de fenômenos presentes no conhecimento em geral ou numa de suas partes.” (PIEADADE, 1983, p.19).

Gênero e espécie também são características classificatórias, e são definidos como:

Gênero é um conjunto de coisas ou ideias que podem ser divididas em dois ou mais grupos ou espécies.

Espécies são os vários grupos resultantes da divisão de um gênero por determinada característica.(PIEDADE, 1983, p.23)

O bibliotecário indiano Ranganathan idealizou o termo faceta, que pode ser empregado como categoria, classe ou gênero, dependendo do sentido para o qual está sendo utilizado (PIEDADE, 1983).

A classificação é feita conforme a necessidade de cada época, pois assim como a sociedade sofre modificações, as classificações também evoluem com o decorrer do tempo. A sociedade classifica, às vezes, de forma diferente ou com nomes diferentes, mas sempre com a mesma finalidade, que é conhecer ou compreender o universo ao qual pertence.

No decorrer da história as primeiras classificações foram criadas pelos filósofos a exemplo de Platão, Aristóteles, Porfírio, Bacon entre outros. Estes classificavam apenas o que conheciam do cosmo e do saber naquele período, formando as Classificações Filosóficas. Posteriormente, com o aumento da produção surgiram as Classificações Bibliográficas, nomes como Dewey, Ranganatham e Cutter foram responsáveis pelo desenvolvimento dos Sistemas de Classificação, em uso até os dias de hoje, que são importantes instrumentos no processo de organização e recuperação da informação em uma biblioteca. A reflexão acerca das classificações filosófica e bibliográfica será feita a seguir .

As bases da Classificação filosófica

O ato de classificar objetiva organizar os seres, as coisas, o conhecimento entre outros, pois Anjos (2008,p.41) enfatiza que:

Tanto o termo quanto o processo de classificação começaram com os gregos, que a usaram com o propósito de tentar organizar os saberes para formar o quadro de conhecimentos da época e tentar conhecer os seres para formar definições, conceitos dos objetos das ciências.

No fim do século VII a.C e durante o século VI a.C, surgiram os primeiros filósofos gregos, conhecidos como físicos ou naturalistas, que pretendiam explicar a totalidade das coisas cosmológicas. Por um bom tempo esses filósofos se empenharam para explicar o surgimento do cosmo, a força originária que age no processo do mesmo, suas fases e os momentos de sua geração.

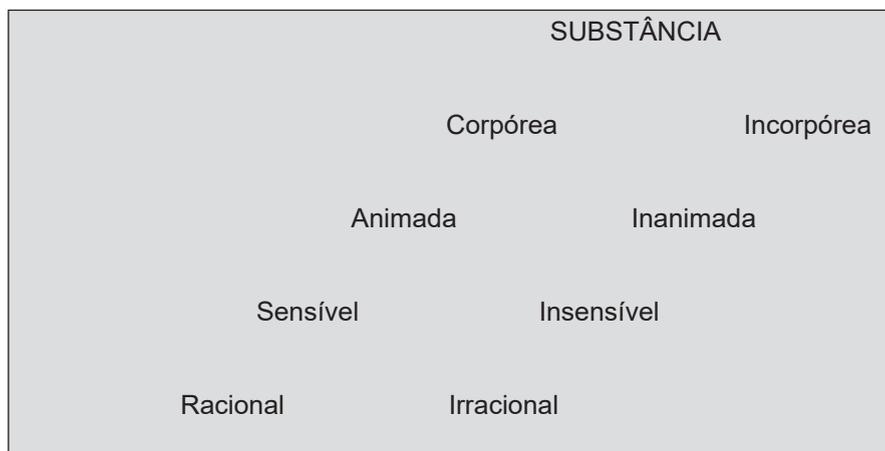
Assim, é possível afirmar que o interesse dos filósofos gregos originou o processo de classificação. É nesse contexto que o filósofo grego Platão destacou-se, pois acreditava que “[...] conhecer significava tanto organizar saberes em sua classe correta, quanto colocar coisas e seres em

sua categoria certa, tudo de acordo com sua permanência ou regularidade [...]” ele foi o primeiro filósofo (do ocidente) a organizar o conhecimento, agrupando-o nas disciplinas: física, ética e lógica. devido sua curiosidade em descobrir a origem do universo e do conhecimento.

Posteriormente, veio seu discípulo Aristóteles que “[...] em sua lógica, mostrava que conhecer consistia em formar e aplicar conceitos, formar juízos e relacioná-los entre si [...]”(ANJOS, 2008, p. 41). Ele propôs uma divisão tritônica e classificou os conhecimentos em teóricos (filosofia e matemática), práticos (ciências exatas e sociais) e produtivos (arte e literatura). Fez essas divisões de acordo com “[...] três distintas operações às quais se dedicam as ciências e que são exercidas pelos homens – pensar, agir e produzir” (ANJOS, 2008, p. 44). A contribuição desses dois filósofos Platão e Aristóteles foi o ponto de partida para a evolução da classificação, e, também serviu de base para outros estudiosos e filósofos, no decorrer da história da humanidade

Porfírio, filósofo grego do século IV, também contribuiu para a classificação, mas deve-se salientar que ele dividiu o conhecimento e “[...] apresentou uma classificação dicotômica, citada em todos os compêndios de classificação bibliográfica, para explicar o processo de divisão das classes, chamada de *Árvore de Porfírio*” (PIEDADE 1983, p. 61). Também chamada de *Árvore de Remée*, pois só foi divulgada pelo francês Pierre de la Remée no século XVI. (Quadro 1).

Quadro 1 – *Árvore de Porfírio*



Fonte: Neves; Kuramoto (2006, p.37)

Porfírio foi o primeiro filósofo a dar um exemplo de classificação binária, Anjos (2008, p. 36) diz que “Nela, os conceitos de Porfírio, são apre-

sentados como subordinando-se logicamente, partindo do geral ao particular, da maior extensão à maior compreensão”, apresentando a relação entre gênero e espécie, como se pode observar no quadro 1.

Na Idade Média, sob, o domínio da igreja Católica, um filósofo se destaca Flávio Magno Aurélio Cassidoro, que, com um pensamento escolástico escreveu livros que serviram de texto para o ensino nas escolas no período de 395-1454. (ANJOS, 2008 ; PIEDADE, 1983,).

Na Classificação de Cassiodoro as ciências foram divididas em dois grupos *Trivium* e *Quatrivium*, (Quadro 2) que compunham 7 disciplinas. Os Estudos Superiores eram compostos por quatro disciplinas. (ANJOS, 2008)

Quadro 2- Classificação Escolástica Romana

TRIVIUM	QUATRIVIUM	ESTUDOS SUPERIORES
Gramática	Geometria	Teologia
Dialética	Aritmética	Metafísica
Retórica	Astronomia	Ética
	Música	História

Fonte: Anjos,(2008, p.46)

Onde as sete primeiras disciplinas (Trivium e Quatrivium) eram preparatórias para os Estudos Superiores. As escolas eclesiásticas pouco a pouco se transformaram nas primeiras universidades. A classificação de Cassidoro por muito tempo foi utilizada, sobretudo, em tais instituições.

Na idade moderna Frances Bacon, filósofo inglês, desde cedo mostrou aversão pela filosofia aristotélica, então ele desenvolveu uma classificação subjetiva, ou seja, voltada para o sujeito em si, que agrupava o conhecimento a partir das “[...] faculdades humanas de MEMÓRIA, IMAGINAÇÃO E RAZÃO” (SOUZA, 2010, p. 15), com destaque para história, poesia e a filosofia. (Quadro 3)

Quadro 3: Classificação de Francis Bacon

FACULDADES MENTAIS		
MEMÓRIA	IMAGINAÇÃO	RAZÃO
Classes		
História (ciências da memória)	Poesia (Ciência da fantasia)	Filosofia (Ciência da Razaõ)
História Natural (Natureza)	Narrativa	Divina (Teologia) Natural

História Civil (Sociedade)	Dramática	(prática, Especulativa) Humana
	Parabólica	(Corpo, alma, social)
Geografia	Belas Artes	Matemática (Ciências Experimentais)

Fonte: Anjos, (2008, p.56)

A classificação de Francis Bacon tornou-se muito importante para a Biblioteconomia e influenciou várias Classificações Bibliográficas, como a de Dewey, Otlet, Cutter, a da Library of Congress (LC) entre outras.

August Comte, famoso filósofo francês, propôs uma classificação que dividia as ciências em abstratas (fundamentais) e concretas (derivadas). “As abstratas estudando as leis gerais, independente dos seres concretos em que se realizam, e as concretas estudando estes seres, considerados em sua complexidade concreta.” (ANJOS, 2008, p.69). Todas estas Classificações Filosóficas serviram de base para as Classificações Bibliográficas as quais serão comentadas a seguir.

Classificação Bibliográfica e seu uso em bibliotecas

A explosão documental, ocasionada pela invenção da imprensa pelo alemão Gutenberg, cópia de documentos. A produção dos cientistas, agora em massa, divulgando deixou para trás o trabalho dos monges copistas, que gastavam anos para fazerem suas pesquisas, através dos livros e periódicos, causaram um volume exorbitante de documentos, e isso fez com que o bibliotecário buscasse organizar toda essa documentação para poder ser encontrada, de acordo com a necessidade de busca de cada pessoa que precisasse de uma determinada informação.

As classificações filosóficas foram criadas com o intuito de descobrir a origem do universo e do conhecimento, mas também foram utilizadas para organizar as ciências da época. Entretanto, “Com o desenvolvimento dos registros do conhecimento e a sua crescente complexidade e especialização, um sistema de classificação filosófico torna-se impróprio para uso em bibliotecas.” (ANJOS, 2008, p.134). Então foram desenvolvidas as classificações bibliográficas ou documentárias, que são utilizadas em bibliotecas, com a finalidade de organização, localização e recuperação dos documentos existentes em sua coleção.

Na Antiguidade os documentos já eram classificados, de maneira mais simples. Calimacus, considerado o primeiro bibliotecário, organizou os livros de acordo com o tipo de escritor: poeta, legisladores, filósofos, historiadores, oradores e escritores de tópicos diversos (PIEADADE, 1983, p.69).

No decorrer do tempo, sob influência das classificações filosóficas, as classificações bibliográficas foram sendo criadas para permitir a organização, localização e recuperação das informações nos acervos das bibliotecas, conforme as necessidades dos leitores. As classificações bibliográficas, segundo Souza, (2010, p.16) podem ser:

- a) Enumerativas, quando arrolam, em números, as categorias em que o universo do conhecimento foi dividido.
- b) Facetadas, quando identificam características comuns a várias categorias de assuntos, organizando-os em facetas.

As classificações enumerativas e facetadas têm a finalidade de descrever o conteúdo do documento através de uma linguagem documental, de maneira fidedigna.

A Classificação é representada por meio da notação que segundo Piedade (1983, p. 38) “[...] é o conjunto de símbolos destinados a representar os termos da classificação, traduzindo em linguagem codificada o assunto dos documentos, e permitindo sua localização [...]”. A notação pode ser pura ou mista. Na notação pura só é utilizado um tipo de símbolo (letra ou número), por exemplo, na CDDir, Direito Eleitoral a representação acontece da seguinte forma:

341.28
 340 Direito
 341.2 Direito constitucional
 341.28 Direito eleitoral

Já na notação mista são utilizados mais de um símbolo (letras, números, sinais gráficos), na CDU, por exemplo, Direito administrativo brasileiro, traduzido para o espanhol, é representado como no exemplo, utilizando o sinal de igualdade (=) para indicar língua abordada no documento.

342.9(81)=134.2
 3 Ciências Sociais
 34 Direito
 342 Direito público
 342.9 Direito Administrativo
 (81) Tabela auxiliar de lugar, 81 representa Brasil
 = 134.2 Tabela auxiliar de Língua, no caso espanhol

É através da notação que os documentos são localizados em meio ao acervo das diferentes bibliotecas.

Outro elemento fundamental das Classificações Bibliográficas é o índice “[...] lista alfabética dos seus termos e respectivos sinônimos, indicando os símbolos de classificação que os representam.” (PIEADADE, 1983, p. 49). Os índices se diferenciam conforme a classificação. Desse modo, eles

podem ser: alfabético, específico, relativo e em cadeia. Conforme Piedade (1983, p. 49- 50)

O índice alfabeto não indica qualquer documento que trata dos assuntos indexados, serve, tão somente, para indicar o símbolo de classificação de determinado assunto

Índice específico é aquele que só relaciona uma entrada para cada assunto.

Índice relativo é aquele que indica para cada fenômeno todos os pontos do sistema em que aparecem os seus vários aspectos.

Índice em cadeia [...] é considerado econômico, [...] possibilitando evitar entradas múltiplas, reduz o número de entradas para cada documento, sem diminuir a eficiência do catálogo ou da biografia.

As principais classificações documentárias são: Classificação Decimal de Dewey, Classificação Decimal Universal, Classificação Expansiva, Classificação da Biblioteca do Congresso, Classificação de Brown, Classificação de Bliss, Classificação de Dois Pontos:

a) Classificação Decimal de Dewey (CDD)

A CDD foi criada por Melville Louis Kossuth Dewey em 1876, que se baseou na classificação de Willian Torrey Harris (a qual foi elaborada para o catálogo da Saint Louis Public School Library) que por sua vez se fundamentou na Classificação de Bacon mas, de forma invertida – razão, imaginação e memória -. Dewey a publicou anonimamente como *A classification and subject index for cataloguing and arranging the books and pamphlets of a librar*. Apenas na 2ª edição publicada em 1885, é que apareceu com o nome do autor Dewey Classification and relative index e só a partir da 16ª edição, ficou denominada como Dewey *Decimal Classification* (DDC), em português Classificação Decimal de Dewey (PIEADDE, 1983, p. 87-88), Ele dividiu o conhecimento em dez classes: uma geral e 9 específicas que abrangesse todos os conhecimentos (Quadro 4).

Quadro 4: Classes principais da CDD

000	Generalidades
100	Filosofia. Psicologia
200	Religião
300	Ciências Sociais
400	Língua
500	Ciências Naturais (Pura)
600	Tecnologia (Ciências Aplicadas)
700	Artes
800	Literatura
900	Geografia, História

Fonte: CDD (2003).

Cada uma dessas classes é subdividida em outras dez e assim por diante, por isso o nome Decimal. Como ilustração, pode-se citar a classificação de um livro de Direito Administrativo

Ex:342.81Direito Administrativo e Constitucional do Brasil

- 300 Ciências Sociais
- 340 Direito
- 342 Direito Administrativo e Constitucional
- 81 Brasil (Tabela 2 .área geografica)

Desde a 18ª edição que a CDD, conta com sete tabelas auxiliares, listadas como segue:

- Tabela 1- Subdivisão padrão,
- Tabela 2- Subdivisão de Área,
- Tabela 3- Subdivisão de Literaturas individuais,
- Tabela 4- Subdivisão de Línguas individuais,
- Tabela 5- Subdivisões de grupos raciais, étnicos e nacionais,
- Tabela 6- Subdivisões de Línguas (idioma),
- Tabela 7- Subdivisão de pessoas.

Na 22ª edição a tabela 7 foi incluída na tabela 2, passando a não ser mais usada. Com isso na Tabela 2 consta Geografia, períodos históricos e pessoas. (CDD, 2003)

A edição mais atual é a 23ª, e, assim como nas quatro ultimas edições, também, é composta de quatro volumes: Volume 1: Prefácio, Introdução, Glossário, Tabelas Auxiliares, volume 2: Classes de 000 -599, Volume 3: Classes de 600-999, Volume 4:Índice Relativo

O índice da CDD é geral, alfabético e relativo, Mendes (1995, p. 35) destaca que “[...] é considerado como um dos melhores em estrutura.” Pois aborda o assunto nas diferentes classes que ele se encontra e emprega margens diferentes a subordinação dos assuntos. A notação é considerada pura, pois são utilizados exclusivamente números arábicos e um ponto, apenas para facilitar na leitura.

A CDD é uma das classificações bibliográficas mais utilizada e mais conhecida do mundo todo. Ela vem rompendo as barreiras do tempo e continua atualizada e útil, sendo empregada em bibliotecas públicas e escolares.

b) Classificação Decimal Universal (CDU)

Os advogados belgas Paul Otlet e Henri La Fontaine idealizaram um repositório bibliográfico universal, para organizar uma bibliografia

dos trabalhos publicados em todo o mundo. Para a orientação sobre as entradas de assunto se basearam na CDD e ficaram maravilhados com a riqueza de seu conteúdo. Como não encontraram o que buscavam, pediram permissão ao próprio Dewey para adicionar e alterar algumas coisas, criando então a 1ª edição da Classificação Decimal Universal (CDU, 2007, p. 8).

A CDU foi a primeira Classificação Bibliográfica que não foi criada com o propósito de organizar documentos nas estantes, mais sim para ser, segundo Anjos (2008, p.1176) “[...] uma ferramenta eficiente de recuperação de informação” eles queriam facilitar o acesso à informação.

Existem cinco edições diferente da CDU que são segundo Piedade (1983, p. 121) “Desenvolvidas, Médias, Abreviadas, Condensadas e Especiais” Cada uma tem suas diferenças que não serão abordadas , pois não é o foco desse estudo.

O índice da CDU é geral e relativo, a notação é mista, pois é composta por números arábicos, sinais gráficos e letras ou palavras. Tem dez classes principais representadas com apenas um algarismo (diferente da CDD que possui três), sendo que sua quarta classe é vaga.(Quadro 5.)

Quadro 5: Classes Principais da CDU

0	Generalidades
1	Filosofia
2	Religião
3	Ciências Sociais
4	Vaga
5	Ciências Naturais
6	Ciencias aplicadas
7	Artes
8	Língua. Literatura
9	Geografia. História

Fonte: CDU (2007)

A estrutura da CDU é formada também por tabelas auxiliares, (Quadro 6) elas são de dois tipos: O auxiliares comuns e auxiliares especiais.

Os auxiliares comuns, além de proporcionarem um meio de expressar inter-relações entre assuntos, indicam característica geralmente repetitivas (isto é aquelas que são aplicáveis em todas as tabelas principais). Os auxiliares especiais indicam características que se repetem em determinado lugares (isto é aquelas que são aplicáveis a um número limitado de tabelas principais) (CDU, 2007, p.18).

Os auxiliares comuns tem dois tipos de símbolos: os sinais e as subdivisões, os sinais são o sinal da adição, a barra oblíqua, os dois pontos duplos, entre outros. Já as subdivisões comuns, são tabelas numéricas, onde os conceitos são enumerados de acordo com a hierarquia determinada (CDU, 2007, p.19).

Quadro 6: Tabelas auxiliares CDU

TABELAS	SIMBOLOS	
1a	+	Coordenação ou adição
	/	Extensão consecutiva
1b	:	Relação simples
	[]	Subagrupamento
	::	Ordenação
1c	=	Língua
1d	(0...)	Forma
1e	(1/9)	Lugar
1f	(=...)	Raça
1g	"..."	Tempo
1h	*	Asterisco
1k	A/Z	Ordem alfabética
	-02	Propriedade
	-03	Materiais
	-04	Relações, processos e operações
	-05	Pessoas
AUXILIARES ESPECIAIS		
	-1/-9	Análítica de traço
	.01/.09	Análítica de ponto
	00/09	Análítica de apóstrofos

Fonte: CDU (2007)

A CDU é mais utilizada em biblioteca universitária e especializada. Devido sua flexibilidade para representar um assunto, a utilização de classificação nas bibliotecas é essencial, pois facilita na busca e recuperação da informação. No exemplo, Constituições brasileiras, no século XIX, mostra o uso das tabelas auxiliares de lugar, tempo.

Ex: 342.4(81)"18" Constituições no Brasil do séc.XIX

342.4 Constituições

- (81) Brasil (Tabela de lugar)
 “18” Século XIX (Tabela de tempo)

Como se pode observar no exemplo acima, os símbolos obedecem a ordem de citação, que é quando “os elementos são combinados para formar um número composto”. (CDU, 2007, p.22). Mas a ordem de arquivamento é utilizada “[...] no arquivamento das fichas no catálogo sistemático, na apresentação de bibliografias e arrumação dos livros na estante” (SOUZA, 2010, p,65).

c) Classificação Expansiva

A Expansive Classification (EC) foi criada por Charles Ammi Cutter em 1891 (baseada na classificação de Bacon, considerado “[...] um dos maiores bibliotecário norte americano” (PIEIDADE,1983,p.146), por não estar satisfeito com a classificação criada por Dewey (CDD), devido a notação decimal adotada.

A EC consiste em sete tabelas classificatórias separadas, “a primeira muito geral, a segunda menos geral, a terceira ainda menos e assim sucessivamente, até a sétima classificação que é bem detalhada.” (PIEIDADE, 1983, p.146). O autor buscava uma classificação que acompanhasse o crescimento do acervo, que fosse utilizada em qualquer tamanho ou a complexidade da coleção, e que possibilitasse o ajustamento do sistema conforme o desenvolvimento do acervo da biblioteca. A sétima tabela não chegou a ser concluída, já que Cutter faleceu antes de finalizá-la.

A notação mista é constituída de letras maiúsculas, algarismos arábicos e ponto. A EC exerceu grande influência na criação da Classificação da Biblioteca do Congresso dos Estados Unidos. Cutter também desenvolveu uma tabela por autor, muito utilizada nas bibliotecas brasileiras.

d) Classificação da Library of Congress (LC)

A LC ou a Biblioteca do Congresso dos Estados Unidos foi fundada em 1800, sua coleção era de 740 livros que primeiramente eram ordenados de acordo com seu tamanho, Em 1812 era organizado seguindo o sistema empregado pela Benjamin Franklin’s Library Company of Philadelphia, que já possuía 18 classes e baseada na classificação de Bacon, usada na Encyclopédie de Diderot e d’Almbert.

Um incêndio que foi provocado pelos britânicos em 1814, no edifício do Congresso, destruiu uma grande parte do acervo. O Congresso em 1815 adquire o acervo particular do ex-presidente Thomas Jefferson. Tal acervo já chegou classificado por um sistema criado pelo próprio Je-

fferson, com 6.487 volumes. Este sistema foi utilizado por um bom tempo, mas com a construção de um novo prédio e com a coleção atingindo cerca de 1 milhão de exemplares decidiram buscar um novo sistema de classificação (PIEDADE, 1983 ; ANJOS,2008)

Examinaram os sistemas da Classificação Decimal de Dewey, a Expansive Classification de Cutter e o sistema de Halle de Hartwing (utilizado na biblioteca da universidade alemã de Halle), então chegaram à conclusão de que iriam criar uma classificação própria, tendo como base a Expansive Classification de Cutter.

Criaram então a Classificação da Biblioteca do Congresso (LCC), que é unicamente para o acervo da LC. Sendo que cada classe foi desenvolvida conforme a coleção. Mas, segundo Anjos (2008, p. 187) "A fragilidade do esquema está na carência de alguma conexão maior entre as classes; não há facetas comuns no índice geral."

A LCC divide o conhecimento em 20 classes, a notação é mista, sendo utilizadas letras maiúsculas, algarismos arábicos de 1 a 9999 e um sinal gráfico: o ponto. Outras bibliotecas nos Estados Unidos e em vários países estão adotando a Classificação do Congresso.

e) Classificação de Brown

James Duff Brown (1861-1914) criou em 1906 com a colaboração de John Henry Quinn, a Subject Classification (Classificação de assunto). Ele foi um dos primeiros ingleses a escrever livros sobre biblioteconomia e o criador do único sistema de classificação geral desse país, devido a reclamações da Classificação de Dewey, acusada de parcialidade em relação aos assuntos americanos. (PIEDADE, 1983, p.173).Brown acreditava que todo conhecimento envolvia dois fatores- o material e a finalidade – combinando ambas formam os assuntos-, chegando a conclusão de que o conhecimento deriva de quatro grupos que são matéria, vida, razão e registro no qual, "Primeiro surgem a Matéria e a Força, que geraram a Vida, esta produziu a Razão, que deu origem ao Registro dos fatos (PIEDADE, 1983, p.174).

A notação é mista constituída de letras maiúsculas , algarismos arábicos, letras minúsculas e sinais gráficos. O sistema está em desuso por falta de atualização, sendo que a ultima foi feita em 1939.

f) Classificação de Bliss

Henry Evelyn Bliss nasceu na cidade de Nova Iorque em 1870, estudou no College of the City of New York, lá exerceu o cargo de bibliotecá-

rio de 1891 a 1940, se dedicou ao estudo da classificação em bibliotecas, criando em 1935 a Classificação Bibliográfica (BC), que é apontada “[...] como um sistema que apresenta um dos melhores desenvolvimentos de classes encontrado em classificação bibliográfica” (PIE DADE, 1983, p. 182)

Ele dividiu o conhecimento de acordo com o conceito educacional da época, “No entender de Bliss, esta ordem é relativamente estável e tende a se tornar mais estável à proporção que o conhecimento se torna mais estabelecido e detalhado.” (PIE DADE, 1983, p.183)

Ele dividiu o conhecimento em quatro grupos: Filosofia, Ciências, História, Tecnologia e Artes, e cada grupo é subdividido em classes. Sua notação é mista, são utilizadas letras maiúsculas e minúsculas, algarismos arábicos, sinais gráficos e também são utilizados o hífen e a vírgula para a separação de tabelas auxiliares.

Ela ganhou uma nova edição feita por Jack Mills, chamada BC2, sendo publicado o primeiro volume em 1977, mas a publicação está atualmente sobre os cuidados de K. G Saur (ANJOS,2008, p.216). Esse sistema ainda é utilizada em algumas bibliotecas do Reino Unido.

g) Classificação de Dois Pontos

O indiano Shiyali Ramamrita Ranganathan, professor de matemática, foi nomeado em 1924 como bibliotecário da *Madras University Library*. Para que adquirese experiência foi mandado para a Inglaterra, a fim de estudar na *School of Librarianship* da Universidade de Londres, onde leu muito sobre Biblioteconomia e se interessou especialmente por classificação e administração de bibliotecas. Ele teve contato com as classificações mais conhecidas como CDD, CDU, CUTTER. Estudou-as e analisou-as.

Regressou à Índia em 1925, para assumir o posto de diretor da *Madras University Library*. Em 1928 ministrou o primeiro curso de Biblioteconomia do seu país, e posteriormente foi trabalhar em *Benares Hindu University*, como chefe da biblioteca e professor de Biblioteconomia (PIE DADE, 1983, p.192-193).

Criou em 1933 a Colon Classification ou Classificação de Dois Pontos, que segundo Souza (2010, p.20) “[...] é usada para relacionar as características, dos diversos assuntos. Sua principal característica é a subdivisão dos assuntos em facetas e focos.” Uma classificação intuitiva, buscando os diferentes aspectos de um assunto ou facetas. Lima (2004, p.68), enfatiza que “Todas as facetas são consideradas como manifestações de uma das cinco categorias fundamentais que são: Personalidade, Matéria, Energia, Espaço, Tempo (PMSET).”

Sua notação é mista sendo utilizada letras maiúsculas e minúsculas, números arábicos, letras gregas e sinais gráficos. Ainda hoje a Classificação de Ranganathan é utilizada nas bibliotecas da Índia.

Em síntese, a base das classificações bibliográficas são as classificações filosóficas, criadas por grandes filósofos como Platão, Aristóteles, Porfírio, Cassiodoro, Bacon entre outros com a finalidade de descobrir a origem do cosmos.

Em biblioteconomia, as classificações bibliográficas foram desenvolvidas para organizar o conhecimento e, entre as mais conhecidas destacam-se: CDD, CDU, EC, LC, BROWN, BLISSE DOIS PONTOS; a estrutura é composta por tabelas que podem ser principais e auxiliares, índice e notação.

O uso das classificações nas bibliotecas são de grande importância para a organização e recuperação da informação, pois é através delas que o documento é classificado de acordo com o assunto, possibilitando a guarda e localização de maneira eficaz.

UM OLHAR SOBRE A CLASSIFICAÇÃO DECIMAL DE DIREITO (CDDir): Histórico e estrutura

O bibliotecário tem uma formação bem generalista e pode atuar em diversas áreas, como hospitais, empresas, escolas, tribunais entre outros. E em cada ambiente ele encontrará documentos específicos, que requeiram para seu tratamento, conhecimentos básicos da área, e assim atender a necessidade do leitor ou pesquisador. Dependendo da área, o profissional pode desenvolver instrumentos que auxiliem na organização e busca da informação, a CDDir é um exemplo.

A Classificação Decimal de Direito (CDDir), foi criada pela bibliotecária Dóris de Queiroz Carvalho, em 1948, na ocasião funcionária da Biblioteca do Ministério da Fazenda. Ela chegou à conclusão que a CDD (que era utilizada para classificar os documentos desta biblioteca), não estava sendo suficiente para suprir as demandas na área jurídica. Então resolveu detalhar a subclasse 340, derivada da classe 300 das Ciências Sociais da Classificação de Dewey, (quadro 4) que é referente ao Direito. (Carvalho, 2002, p.3).

Surge assim, a Classificação Decimal de Direito, baseada na CDD. Ela foi reeditada em 1953, e devido a vários pedidos, em 1977 foi editada a “[...] a terceira edição publicada pela Divisão de Documentação do Ministério da Fazenda. Em 2002 foi lançada a quarta edição [...]”

A CDDir tem em sua estrutura 4 classes principais (quadro 7), números de forma e um Índice. As Classes principais estão divididas em “Direito Público e Direito Privado, com exclusão do Direito Canônico e do Direito Romano, que ficaram situados fora desses dois grupos” (CARVALHO, 2002, p.7), os dois últimos, não se acomodam na “divisão clássica do direito” (privado e público) já existentes há muitos anos e são chamados por Passos (2009, p.3) de “outros ramos do direito” por isso foram classificados com sua própria classe.

Quadro 7:Classes principais da CDDir

340	Direito
341	Direito Público
342	Direito Privado
343	Direito Canônico
344	Direito Romano

Fonte: Carvalho, 2002, p.7.

Assim como na classificação de Dewey, as classes da CDDir também se subdividem em subclasses, conforme a especificidade do assunto.

Os números da tabela de forma¹ (Quadro 8) “[...] são adotados para agrupar, dentro de uma classe geral ou subdivisão, o material bibliográfico que apresenta certas características especiais na maneira em que o assunto é tratado,” ou seja, tem a função das tabelas auxiliares, adotadas em outras classificações.

Quadro 8:Números de forma da CDDir

01	Teoria, filosofia
02	Compêndios
03	Dicionários e enciclopédias
04	Discursos, ensaios, conferências
05	Periódicos
06	Sociedades
07	Estudos e ensino
08	Coleções, poligrafia
09	A história de um assunto ou sua divisão por país

Fonte: Carvalho, 2002, p.8.

¹ Forma serve para representar a forma sob a qual está apresentado o documento ou assunto (PIEDADE, 1983,p. 139)

Exemplificando, o uso do número de forma, um periódico de Direito, ficaria classificado pela CDDir desta maneira:

Ex: 340.05 Periódico de Direito.

Sua notação assim como a CDD, é pura, pois só é utilizado um tipo de símbolo que são os algarismos arábicos e utiliza um ponto somente para facilitar na leitura, não tendo qualquer valor na notação. Como pode ser visto abaixo.

Ex: 342.299 Direito aeronáuticos.

O Índice da Classificação de Direito é alfabético, e assim como os índices de outras classificações, serve de auxílio para a consulta às tabelas. O bibliotecário precisa “[...] consultar sempre o índice antes de classificar qualquer publicação, [...] Porém, nunca deve ser feita a classificação diretamente pelo índice, sem consulta às tabelas principais, o que poderá induzir o classificador a graves erros.” (CARVALHO, 2002, p.9).

A CDDir é um importante instrumento na organização e disseminação da informação, servindo de apoio aos diversos bibliotecários que trabalham em um campo tão específico que é o jurídico.

A CDDir está disponível tanto em meio eletrônico por meio do site www.planalto.gov.br, gratuitamente, como impressa. Qualquer pessoa jurídica ou física pode ter acesso.

A CLASSIFICAÇÃO BIBLIOGRÁFICA E A ORGANIZAÇÃO E RECUPERAÇÃO DA INFORMAÇÃO NA ÁREA JURÍDICA

A biblioteca é uma das instituições que tem como função a guarda da memória das organizações, cidades, países entre outros, e a organização e recuperação do conhecimento que também são de sua responsabilidade. A organização do conhecimento serve “[...] não apenas para identificar itens de informação específicos e definidos de forma precisa, mas também para demonstrar a completa gama de assuntos disponíveis na biblioteca e suas relações entre si.” (LANGRIDGE, 2006, p.19). Ela é feita por meio de várias atividades desenvolvidas pelo bibliotecário tornando possível a recuperação da informação, seja para pesquisas, estudos ou lazer.

Os serviços da Biblioteca são constituídos por vários subsistemas, dentre eles se destacam dois, que são o subsistema de entrada dividido em: desenvolvimento da coleção, tratamento da informação e armazenagem. E o de saída, dividido em: análise/ negociação de questões, estratégia de busca, busca e disseminação. Entre os processos, tem um que

é essencial tanto na entrada da informação quanto na saída, que é a representação documental (DIAS; NEVES, 2007, p.15), como o nome já diz “a representação” é o ato de representar algo. Neste caso, o “algo” a ser representado é o documento, e por meio da representação ele pode ser organizado e recuperado.

A palavra representação é um termo polissêmico, uma vez que possui diferentes significados e é adotada por diversos ramos do conhecimento. A Linguística e a Semiótica, por exemplo, usam a língua para representar algo, a Comunicação representa a **mediação entre emissor e receptor**, a Biblioteconomia / Ciência da Informação usam a representação para caracteriza os documentos. (DODEBEI, 2002).

A Representação documental pode ser descritivas ou temáticas ambas têm a mesma função que é disponibilizar dados sobre o documento, sejam intelectuais ou descritivos e/ou físicos. Através dessas atividades é que a recuperação da informação ocorre, feita por um profissional capacitado que se preocupa com a disponibilização da informação e o atendimento ao leitor.

a) A representação descritiva também chamada de descrição física tem a função de descrever as características físicas dos documentos. Dias e Naves (2007, p. 19) dizem que “O processo de descrição física compreende [...] um exame do documento com o objetivo de identificar certos elementos nele constantes e que vão servir para identificá-lo.” Os processos fundamentais desenvolvidos por ela são a catalogação e a referência.

A Catalogação “[...] consiste em um conjunto de informações que simbolizam um registro do conhecimento”. Com as informações são feitas fichas chamadas fichas catalográficas. No Brasil foi adotado O Código de Catalogação Anglo Americano (AACR) para padronizar as informações e organizá-las. “À atividade de descrição dos aspectos físicos costuma se dar o nome de catalogação descritiva, e a do conteúdo, catalogação por assunto.” (DIAS; NAVES, 2007, p.16), as entradas de autor são feitas com base no AACR enquanto que as de conteúdo são através de cabeçalho de assunto ou por sistemas de classificações.

O “conjunto de fichas que representam as publicações da coleção da biblioteca constituem os catálogos, que permite a sua recuperação sob diferentes aspectos” (CRUZ, 2004, p.25), os catálogos também estão sendo feitos em formato eletrônico, que é semelhante ao tradicional.

A Referência assim como a catalogação, serve para descrever o documento fisicamente, a Norma Brasileira Registrada (NBR) 6023 da Associação Brasileira de Normas Técnicas (ABNT) no Brasil é responsável pela

determinação e identificação dos elementos que formam a referência, como autor, título, editora entre outros.

A importância da referência é enfatizada por Nahuz e Ferreira (2007, p.121) como essencial para o pesquisador, pois “A referência dos documentos utilizados em um trabalho é um dos recursos de representação desses documentos que permitem sua identificação e localização facilitando aos estudiosos/ pesquisadores chegar facilmente a eles[...]” essa representação orienta o leitor e mostra novas opções de leituras que pode ser útil a sua pesquisa.

b) A Representação temática tem a função “[...] de identificar o assunto ou assuntos do documento.” (DIAS; NAVES, 2007, p.20), ou seja, identificar seu conteúdo intelectual, através da leitura. Os principais processos desenvolvidos por ela são a indexação e o resumo.

A Indexação é uma das atividades central desenvolvida na biblioteca, na qual o documento é representado por descritores. Para Guinchat e Menou (1994, p. 175) “É a operação pela qual escolhe-se os termos mais apropriados para descrever o conteúdo de um documento.” Está presente tanto na entrada, como na saída da informação, de uma biblioteca. A indexação dos documentos é, feita na entrada pelo bibliotecário e na saída é indexada a “questão” solicitada pelo usuário. É uma tarefa desenvolvida pelo bibliotecário, que precisa ter “[...] conhecimento prévio linguístico, textual, de mundo [...]” (FUJITA; RUBI, 2006, p. 2), que busca extrair os termos de acordo com a política adotada na instituição na qual trabalha (especializada, universitária entre outras).

A leitura é uma atividade essencial, no processo de indexação que exige muita atenção, pois é através desta que são extraídos os termos, que identificarão o assunto contido no documento. A leitura documental é diferenciada, pois “[...] consiste em uma leitura direcionada para certas partes do documento onde vai-se encontrar elementos especialmente importantes para a identificação do assunto ou assuntos do documento.” (DIAS; NAVES, 2007, p. 53), esta leitura também chamada de leitura técnica, concentra-se no título, sumário, resumo, introdução, título de cada capítulo. A leitura do documento completo seria quase impossível, devido ao tempo para a disponibilização do documento e a escassez de pessoal capacitado para tal atividade.

Depois de extrair os termos é preciso traduzi-los para uma linguagem de indexação, Dias e Naves (2007, p. 23) dizem que

Existem dois tipos principais de linguagens de indexação: Linguagens alfabética e linguagens simbólicas. Os principais tipos de linguagens alfabética que existem são as listas de cabeçalho de assunto e os tesau-

ros. [...] quanto às linguagens simbólicas, estas se limitam praticamente aos sistemas de classificação bibliográficas, [...]

É com o auxílio das linguagens de indexação que ocorre uma padronização para representar os assuntos identificados nos documentos, possibilitando, com isso, o acesso à informação.

A análise conceitual feita através da leitura para a atribuição de descritores e a tradução dos termos para uma linguagem documentária são as duas etapas que compõem a indexação. Ela também já está sendo feita automaticamente, ou seja, “[...] consiste em fazer o computador reconhecer palavras que aparecem no título, no resumo do documento, ou no próprio texto. Os termos reconhecidos são incorporados em um arquivo de pesquisa e servem para recuperar o documento.” (GUINCHAT; MENO, 1994, p.182). Mas mesmo com toda tecnologia, a indexação automática necessita do acompanhamento de um profissional, para ser eficaz.

As representações temáticas e descritivas são tarefas que requerem do bibliotecário muita atenção, pois é por meio das informações corretas que os documentos chegam até seu respectivo leitor.

É com base em tal processo que, a seguir, será feita uma análise da classificação, etapa da representação temática na área jurídica, especificamente no domínio dos Direitos Administrativos e Constitucional.

Caracterizando o domínio do Direito Administrativo e Direito Constitucional

A área escolhida para a análise foi o Direito, especificamente dois ramos o Direito Administrativo e Constitucional que serão o objeto desse estudo. O primeiro é conceituado como “Conjunto harmônico de princípios Jurídicos que regem os órgãos, os agentes e as atividades públicas tendentes a realizar concreta, direta e imediatamente os fins desejados pelo Estado” (MEIRELLES, 2004, p.40), ou seja, é voltado para a administração pública. Enquanto o segundo, “É a parcela da ciência jurídica destinada ao estudo da constituição, em suas diversas acepção ou classificação [...]” (ARAUJO, 2011, p.25), a constituição é a lei maior de um país.

Tal área (Direito) foi levada em consideração, por se tratar de uma área extremamente importante para a sociedade, pois é a área voltada para a conduta do cidadão. Logo, a demanda por informação neste domínio não se limita aos alunos, professores e juristas, mas se estende à comunidade em geral, por estarem presentes em conteúdos de concursos públicos.

Primeiramente, busca-se conceituar a área em questão, para ajudar na análise por isso, foram consultados dicionário e enciclopédias (Quadro 9 e 10), assim como a literatura científica dos termos propostos para ter uma compreensão melhor, de como estão organizados e representados na CDU e CDDir.

Quadro 9: Definição de termos em Dicionário

TERMOS	DEFINIÇÃO
Direito	Ciência das normas obrigatórias que disciplinam as relações dos homens em sociedade.
Administrativo	Relativo à administração.
Constitucional	Relativo ou pertencente à constituição.

Fonte: Ferreira, (2009).

No dicionário pesquisado, os termos Direito Administrativo e Direito Constitucional não foram encontrados, entretanto encontrou-se termos Administrativo e Constitucional que podem dar uma direção do significado dos termos procurados. Com base nas definições encontradas, diz-se que, Direito Administrativo, são normas e regras voltadas para a administração tanto pública quanto privada, e Direito Constitucional é o direito voltado para a Constituição que é a lei maior que rege uma nação.

Quadro 10: Definição de termos da Enciclopédia

TERMOS	DEFINIÇÃO
Direito	Complexo de leis ou normas que regem as relações entre os homens.
Direito Administrativo	Ramo do direito público que tem por objetivo o funcionamento dos serviços públicos e sua relação com os particulares.
Direito Constitucional	Ramo do Direito Público interno dedicado à análise e interpretação das normas constitucionais.

Fonte: Enciclopédia Britânica (2013)

Embora os sentidos dos termos Direito, Direito Administrativo e Direito Constitucional, sejam registrados em linguagem natural nos dicionários de língua e em enciclopédia como observados nos quadros 9 e 10, estas fontes proporcionam o entendimento e a compreensão de tais conceitos. Porém, vale ressaltar que as possibilidades de expressão em linguagem natural interferem na representação e, conseqüentemente, na recuperação da informação, na medida em que restringe as possibilidades de coincidências entre a linguagem adotada pelo autor e sua representação

O Direito é essencial dentro de uma sociedade, pois "É o conjunto de regras de conduta coativamente imposta pelo estado." (MEIRELLES, 2004,

p.37). Ele é tradicionalmente dividido em Direito Público e Direito privado.

O Direito Público tem por objetivo principal a regularização dos interesses estatais e sociais, só alcançando as condutas individuais de forma indireta. Integram esse ramo o Direito Constitucional, o Direito Administrativo, o Direito Tributário, o Direito penal, entre outros. [...] O Direito Privado tem como preocupação principal a regularização dos interesses individuais, para possibilitar o convívio das pessoas em sociedade e uma harmônica fruição de seus bens [...]. O Direito Comercial e o Direito Civil são os integrantes típicos do Direito Privado. (ALEXANDRINO; PAULO, 2006, p.1).

A divisão clássica do Direito em Direito Privado e Direito Público teve sua origem no Direito Romano. Essas duas grandes divisões dos assuntos jurídicos, que embora continuem válidas, atualmente é motivo de polêmica entre os especialistas da área jurídica, pois do ponto de vista científico “têm, sido falhos, insuficientes e obscuros os critérios distintivos” entre tais ramos jurídicos. (CARVALHO apud PASSOS; BARROS (2009, p.1)

Entretanto, Narukawa, et al (2012, p.4) por meio das orientações encontradas na literatura e das orientações obtidas em consenso dos docentes, estabeleceram a seguinte estruturação dos assuntos da área jurídica:

- 1 DIREITO
 - 1.2 DIREITO PÚBLICO
 - 1.2.1 Direito Administrativo
 - 1.2.2 Direito Constitucional
 - 1.2.3 Direito Econômico
 - 1.2.4 Direito Financeiro
 - 1.2.5 Direito Penal
 - 1.2.6 Direito Processual
 - 1.2.6.1 Direito Processual Civil
 - 1.2.6.2 Direito Processual Penal
 - 1.2.7 Direito da Seguridade Social
 - 1.2.8. Direito Tributário
 - 1.2.9 Direito Urbanístico
 - 1.3 DIREITO PRIVADO
 - 1.3.1 Direito Civil
 - 1.3.2 Direito Comercial
 - 1.3.3 Direito Romano
 - 1.3.4 Direito do Trabalho
 - 1.4 DIREITOS ESPECIAIS
 - 1.4.1 Direito Ambiental

- 1.4.2 Direito Canônico
- 1.4.3 Direito do Consumidor
- 1.5 DIREITO INTERNACIONAL
- 1.4.3 Direito do Consumidor
- 1.5 DIREITO INTERNACIONAL
- 1.5.1 Direito Internacional Público
- 1.5.2 Direito Internacional Privado
- 1.6 FILOSOFIA DO DIREITO
- 1.7 HISTÓRIA DO DIREITO
- 1.8 SOCIOLOGIA DO DIREITO
- 1.9 TEORIA GERAL DO DIREITO

Com base nessa hierarquização dos conceitos, é possível ter uma visão panorâmica da área jurídica.

Desvelando a Classificação dos Direitos Administrativos e Constitucional na CDU e CDDir

As representações do Direito Administrativo e Direito Constitucional na CDU, estão localizadas nas subclasses 342 e 342.9 as quais são subordinadas à classe 34 que representa o Direito, que, por sua vez, está subordinada à grande classe 3 de Ciências Sociais (Quadro11).

Quadro 11: Classe de Direito da CDU.

3	Ciências Sociais
34	Direito
342	Direito público. Direito Constitucional
342.4	Constituições. Assembleias legais
342.41	Bases da Constituição
342.9	Direito Administrativo
342.92	Direito Administrativo geral
342.94	Atos legais. Atos administrativos
342.95	Direito administrativo especializado
342.951	Direito administrativo de áreas especializadas
342.97	Direito administrativo das organizações
342.98	Direito do pessoal administrativo, dos funcionários públicos.

Fonte: CDU, 2008

Na CDDir, o Direito Administrativo e Direito Constitucional, estão subordinados à classe 341 de Direito Público, que é subordinada à classe 340 de Direito. (quadro 12). Observando os quadros 11 e 12, pode-se afirmar que a CDU aborda somente assuntos específicos dos Direitos Administrativos e Constitucional, enquanto a CDDir engloba os outros domínios que lhes são subordinados. Isto ocorre devido a CDDir ter sido criada especificamente para classificar a massa documental da área jurídica, por um profissional da área, que conhecia a necessidade de organização e recuperação da informação.

Quadro 12: Classe de Direito na CDDir.

340	DIREITO
341	DIREITO PÚBLICO
341.2	DIREITO CONSTITUCIONAL
341.253	Direito parlamentar. Paramentos e assembléias
341.27	Direitos Fundamentais. Direitos primordiais
341.28	Direito eleitoral. Sistemas eleitorais
341.3	DIREITO ADMINISTRATIVO
341.347	Direito Ambiental
341.3479	Direito ambiental internacional
341.378	Direito Econômico
341.38	Direito Financeiro
341.39	Direito tributário

Fonte: Carvalho, 2002

O Interesse nestes dois ramos do Direito é por serem utilizados não apenas pelos profissionais da área jurídica, mas sim por pessoas de várias áreas e diferentes graus de escolaridade, e por estarem entre os assuntos que fazem parte do conteúdo programático dos concursos públicos. Como exemplo os quadros 13 e 14 mostram o conteúdo de Direito Administrativo e Constitucional, para o nível Fundamental, Médio e Superior, do concurso da Universidade Federal do Maranhão.

Quadro 13: Conteúdo programático do Concurso da UFMA 2012

Legislação: 1. Direito Administrativo: Lei 8.112 de 11/12/90 que dispõe sobre o Regime Jurídico dos Servidores Públicos da União, das Autarquias e dá outras providências; Lei 8.666 de 21/06/93 que regulamenta o art. 37, XXI da Constituição Federal e institui normas para licitações e contratos da Administração Pública e dá outras providências. 2. Constituição Federal: Título II - Dos Direitos e Garantias Fundamentais; Título III - Da Organização do Estado (Capítulo VII - Da Administração Pública).

Fonte: <http://www.concursos.ufma.br/editais/0212.pdf>

Quadro14: Conteúdo de Direito do concurso do MPU 2013

NOÇÕES DE DIREITO CONSTITUCIONAL E ADMINISTRATIVO: 1 Constituição. 1.1 Conceito, classificações, princípios fundamentais. 2 Direitos e garantias fundamentais. 2.1 Direitos e deveres individuais e coletivos, direitos sociais, nacionalidade, cidadania, direitos políticos, partidos políticos. 3 Organização político-administrativa. 3.1 União, estados, Distrito Federal, municípios e territórios. 4 Poder legislativo. 4.1 Congresso nacional, câmara dos deputados, senado federal, deputados e senadores. 5 Poder executivo. 5.1 Atribuições do presidente da República e dos ministros de Estado. 6 Poder judiciário. 6.1 Disposições gerais. 6.2 Órgãos do poder judiciário. 7 Funções essenciais à justiça. 7.1 Ministério público, advocacia e defensoria públicas. 8 Agentes públicos. 8.1 Espécies e classificação. 8.2 Cargo, emprego e função públicos. 8.3 Regime jurídico do Servidor Público. 9 Poderes administrativos. 9.1 Hierárquico, disciplinar, regulamentar e de polícia. 9.2 Uso e abuso do poder. 10 Licitação. 10.1 Princípios, dispensa e inexigibilidade. 10.2 Modalidades. 10.3 Lei nº 8.666/1993. 11 Controle e responsabilização da administração. 11.1 Controles administrativo, judicial e legislativo. 11.2 Responsabilidade civil do Estado.

Fonte: http://www.cespe.unb.br/concursos/mpu_13_2/arquivos.

Os quadros 13 e 14, expõem o conteúdo programático referente ao Direito Administrativo e Constitucional, proposto tanto pela UFMA como pelo Ministério Público da União, para o preenchimento de vagas para cargos de níveis Médio e Superior. Mostra também a relação entre os termos que constituem estes dois ramos sob a ótica das instituições envolvidas.

Apartir dos conteúdos sumarizados nos quadros 13 e 14 foram escolhidos cinco termos para compor o *corpus* da análise, a saber: Licitação e contratos, Regime Jurídico do Servidor Público, Direitos e garantias fundamentais, Poderes do Estado e Atos administrativos e esboçada uma caracterização conceitual básica nos dicionários, conforme quadros 15 e 16.

Quadro: 15 Definições de termos do Dicionário

TERMOS	DEFINIÇÕES
Licitação	Procedimento pelo qual a administração pública seleciona a proposta mais vantajosa, quando compra bens e serviços ou faz outras transações.
Contrato	Acordo entre duas ou mais pessoas que transferem entre si algum direito ou se sujeitam a alguma obrigação.
Regimento	Normas impostas ou consentidas; disciplina. Conjunto de normas que regem o funcionamento de uma instituição:
Direito	Aquilo que é justo, reto e conforme a lei.
Garantia	Ato ou palavra com que se assegura uma obrigação, uma intenção, segurança.
Poder	Dispor de força ou autoridade:
Estado	O conjunto dos poderes políticos de uma nação; governo:
Ato Administrativo	Documento público em que se exprime decisão de uma autoridade

Fontes: Ferreira, (2009).

Licitação e contrato, direitos e garantias e Poder do Estado foram conceituados separadamente por não terem sido encontrados juntos, mas Pode-se ter uma base dos seus significados através de suas definições. O primeiro termo licitação e contrato são as atividades desenvolvidas pela administração pública para a aquisição de bens ou serviços e o pacto entre os contratantes.

O segundo, como o próprio nome diz, são os direitos e garantias assegurados pelos cidadãos. Já o terceiro é o Poder que o Estado tem sobre a administração do país ou nação.

O termo Regime Jurídico do Servidor Público não foi encontrado, entretanto foi localizado outro enunciado relacionado que deu uma direção ao significado do termo proposto. Logo o significado do termo é um conjunto de normas que conduz o servidor público.

E, por último ato administrativo foi encontrado o termo completo. Trata-se de um documento regido para impor alguma decisão da administração pública.

A seguir, as definições dos enunciados em um dicionário da área, para melhor compreensão.

Quadro:16: Definição de termos em Dicionário Jurídico.

TERMOS	DEFINIÇÃO
Licitação	É o procedimento administrativo formal para contratação de serviços ou aquisição de produtos da Administração pública direta ou indireta.
Contrato	É o acordo recíproco de vontades que tem por fim gerar obrigações recíprocas entre contratantes.
Regime Jurídico do Servidor Público	É o título correspondente à lei 8.112 de 1990, destinada a regular a carreira do Servidor Público brasileiro.
Direitos e garantias fundamentais	É o termo referente a um conjunto de dispositivos contidos na Constituição brasileira de 1988 destinados a estabelecer direitos, garantias e deveres aos cidadãos brasileiros.
Poder	Direito de deliberar, agir, mandar e também dependendo do contexto, a faculdade de exercer a autoridade, a soberania.
Estado	Conjunto das instituições que controlam e administram uma nação.
Ato Administrativo	É toda manifestação unilateral de vontade da administração pública que, agindo nesta qualidade, tenha por fim imediato resguardar, adquirir, modificar, extinguir e declarar direitos ou impor obrigações aos administrados ou a si própria.

Fonte: Santos (2001).

Os termos foram encontrados completos com seus devidos significados, com exceção de Licitação e Contrato e Poderes do Estado, que foram comentados separadamente, onde suas definições seriam basicamente as mesmas, propostas pelo dicionário.

De posse das definições dos dicionários, considerar-se preliminarmente os seguintes conceitos relacionados aos termos escolhidos.

- a) Sobre Licitação pode-se estabelecer os seguintes conceitos:

Procedimentos administrativos

Administração pública

- b) Sobre Contrato:

Acordo recíproco

- c) Regime Jurídico do Servidor Público:

Regularização do servidor público

Comportamento dos servidores

Lei 8.112

- d) Poder do estado

Soberania estatal

- e) Ato administrativo

Manifestação da administração pública

Decisões administrativas

Convém resaltar que a caracterização conceitual foi possível a partir da consulta à literatura especializada. Essa pré-organização dos conceitos fundamentais dos domínios Direito Administrativo e Constitucional, complementa-se com uma análise dos Sistemas de Classificação.

Na literatura científica o termo licitação é conceituado como “[...] procedimento administrativo que, para fins de contratação, objetiva a seleção da melhor proposta entre as apresentadas, seguindo regras objetivas, respeitada a isonomia entre os participantes.” (PRADO, 2009, p. 149). É o modo que a administração pública utiliza para contratar serviços ou produtos de maneira justa entre os interessados. O termo contrato é a forma de acordo entre as duas partes interessadas, Prado (2009, p. 149) o conceitua como “[...] pacto entre duas ou mais pessoas, que se obrigam a cumprir o que foi entre elas combinado sob determinadas condições.”

O Termo Regime Jurídico do Servidor Público, refere-se à Lei 8.112/90 que rege os Servidores Públicos Cíveis da União, ela conduz os direitos e deveres dos servidores públicos no exercício do cargo público. Segundo

o art. 2 desta lei (1990, p.1) servidor público é “[...] a pessoa legalmente investida em cargo público.” Foi encontrado na mesma lei o conceito de cargo público “[...] o conjunto de atribuições e responsabilidades previstas na estrutura organizacional que devem ser cometidas a um servidor.” O Regime jurídico engloba o ambiente das autarquias e fundações públicas federais, direcionando o comportamento dos servidores.

Os termos direitos e garantias fundamentais como o nome já diz, trata-se dos direitos e garantias que o cidadão tem em seu país. O art. 5 da Constituição brasileira (1988, p.15) diz “Todos são iguais perante a lei, sem distinção de qualquer natureza, garantindo-se aos brasileiros e aos estrangeiros residentes no país a inviolabilidade do direito à vida, à liberdade, à segurança e à propriedade [...]”. Tal artigo mostra, de forma clara, a definição dos termos.

O termo Poder do Estado trata da autoridade que o estado exerce

sobre uma nação., na Constituição brasileira o Poder do estado está dividido em três o Legislativo que é exercido pelo Congresso Nacional, o Executivo que é exercido pelo presidente da República, auxiliado pelos Ministros do Estado e o Judiciário exercido pelos juizes e órgãos da mesma natureza (CONSTITUIÇÃO, 1988)

E por último o termo Ato Administrativo são providências que complementam as leis, para o melhor desempenho das atividades públicas, Meirelles (2004, p.76) afirma que

O ato administrativo é toda manifestação unilateral de vontade da Administração Pública que, agindo nessa qualidade, tenha por fim imediato adquirir, resguardar, transferir, modificar, extinguir e declarar direitos ou impor obrigações aos administrados ou a si própria.

O ato administrativo é essencial para os órgãos públicos e para os funcionários, pois é através dele que são feitas as modificações exigidas pela Administração Pública.

a) Licitação e Contratos

Na CDU quando localizada a classe 3, encontram-se os seguintes termos para Licitação e contrato:

347.238.8 Dissolução de propriedade conjunta. Licitação, venda em leilão em um só lote.

351.712.2 Contratação de obras e suprimentos por órgãos públicos. Contratação.

Na CDDir os termos licitações e contratos estão representados da seguinte forma:

341.352 Contratos administrativos

341.3527 Licitações em geral

Pode-se observar que na CDU os termos não se encontram dentro da subclasse de Direito Administrativo, uma vez que o primeiro está subordinado à classe 347, referente ao Direito Civil e o segundo à subclasse 351, referente a Atividades específicas da administração pública da classe 35 Administração Pública enquanto na CDDir ambos estão localizados na subclasse 341.3 de Direito Administrativo.

No SISTEMA DE BUSCA SIGAA DA BC ao ser colocado o termo Licitação e contratos, apareceram vários títulos. Dentre alguns:

Licitação e contratos Administrativo com a classificação 351.712.2.

Licitação e contratos: a lei nº 8.666/93 simplificada com a classificação 351.712.2.

Licitação e contratos: teoria, jurisprudência e mais de 440 questões fundamentais sua classificação é 351.712.2(81). As classificações estão de acordo com a CDU que é o Sistema de Classificação adotado pela instituição, mas não está de acordo com a literatura, pois não estão localizados na subclasse do Direito Administrativo.

Na Biblioteca do TJ foram encontrados os seguintes títulos referentes ao termo:

Licitação e contrato Administrativo com a classificação 341.3527

Licitação com a classificação 341.3527

Podemos afirmar que ambos os documentos estão classificados de acordo com o Sistema Classificatório adotado pela instituição a CDDir, também estão de acordo com a literatura, pois pertencem a classe de Direito Administrativo.

b) Regime Jurídico do Servidor público

Na CDU foram localizados os termos:

342.98 Direito do pessoal administrativo, dos funcionários públicos.

35.08 Pessoal dos órgãos públicos dirigentes. Funcionários. Quadro de pessoal. Serviço público: teoria, organização. Funcionalismo como um todo.

Na CDDir foi encontrado:

341.3313 Servidores regidos pela legislação trabalhista

Observando o primeiro termo na CDU percebe-se que ele é subordinado à classe 342 de Direito Público, Direito constitucional, e o segundo é localizado da Classe 35 Administração Pública, sendo que o correto seria que o termo estivesse na subclasse do Direito Administrativo. Na CDDir o termo está bem representado, por ser um termo do Direito Administrativo.

Ao procurar-se o termo no sistema da Biblioteca Central da UFMA, este não foi encontrado. Então colocou-se o termo Lei 8112 e imediatamente, apareceram vários títulos que foram destacados:

O servidor da Lei nº 8.112/90 no concurso público: 110 questões comentadas com a classificação 35.081.7(81)

Regime Jurídico dos Servidores públicos civis da união, das autarquias e das fundações públicas federais. Com a classificação 35.081.7(81) (094)

Na Biblioteca do TJ foi encontrado o título:

Servidores Públicos Federal: lei 8.112/90 com a classificação 341.330981

Observa-se que na BC da UFMA os documentos estão classificados de acordo com a CDU, mas não de acordo com a literatura, pois, a Lei 8.112 pertence ao Direito Administrativo. Na Biblioteca do TJ o documento está classificado de acordo com a CDDir, e está de acordo com a literatura, pois Regime do Servidor Público é um assunto de Direito Administrativo.

c) Direitos e garantias fundamentais

Na CDU 342.7 Direitos fundamentais. Direitos humanos. Direitos e deveres do cidadão.

Já na CDDir foi encontrado o seguinte termo:

341.27 Direitos Fundamentais. Direitos primordiais. Garantias asseguradas aos cidadãos e associações. Direitos do homem. Liberdades fundamentais. Direitos políticos.

O termo Garantia não foi localizado, enquanto o segundo está localizado

na subclasse do Direito Público, Constitucional, e na CDDir o termo também encontra-se na subclasse do Direito constitucional, onde ambos estão bem representados conforme a literatura.

Na Biblioteca Central os termos mostraram os seguintes títulos:

A (de)limitação dos direitos fundamentais com a classificação 342.7

A Exigibilidade dos direitos fundamentais sociais no estado democrático de direito 342.7.

Na Biblioteca do TJ foram encontrados:

Princípios, direitos e garantias fundamentais: constituição e legislação infraconstitucional que está classificado 341.2

Em ambas as bibliotecas os termos estão de acordo com os Sistemas adotados (CDU e CDDir). E também estão de acordo com a literatura, pois direitos e garantias fundamentais são termos pertencentes ao Direito Constitucional.

d) Poder do Estado

Na CDU foi encontrado o seguinte termo:

342.5 Poder do Estado Sistema e função dos órgãos de governo.

E na CDDir o termo encontrado foi:

341.25 Poderes do Estado

Os termos de ambas as classificações estão dentro da subclasse do Direito Constitucional, estão bem representados, pois é um termo pertencente ao Direito Constitucional.

Na Biblioteca Central foi encontrado o seguinte título:

Da organização do estado, dos poderes e histórico das constituições. Sua classificação é 342

E na Biblioteca do TJ há o mesmo título:

Da organização do estado, dos poderes e histórico das constituições é classificado 341.2

Os poderes do estado: funções atuação e relacionamento com a seguinte classificação 341.25.

Os títulos de ambas as bibliotecas, estão classificados de acordo com os Sistemas de Classificação adotados pelas instituições (CDU e CDDir), e também estão de acordo com a literatura, pois o termo poder do estado, pertence ao Direito Constitucional.

e) Atos Administrativos

O termo é representado na CDU como:

342.94 Atos legais. Atos administrativos

E na CDDir o termo está representado assim:

341.35 . Atos administrativos

Os termos de ambas as classificações estão dentro da subclasse do Direito Constitucional, e estão bem representados, pois é um termo pertencente ao Direito Constitucional.

Na Biblioteca Central foram encontrados os seguintes títulos:

Atos administrativos com a classificação 35.077.2

Atos da administração lesivos ao patrimônio público dos princípios constitucionais da legalidade e moralidade. Com a classificação 35.077.2

Na Biblioteca do TJ foi encontrado o seguinte título:

Teoria dos atos administrativos. Classificado como 341.3501

Limites à razoabilidade nos atos administrativos. Classificado como 341.25.

Nas classificações realizadas pela Biblioteca Central da UFMA, ambos os exemplos, não estão de acordo com a CDU, pois, estão classificados na subclasse de Administração Pública, sendo que o correto seria que o termo estivesse na subclasse 342.94 referente a Atos administrativos. Na classificação da Biblioteca do TJ, o primeiro título está classificado de acordo com o número referente a atos administrativos na CDDir, enquanto o segundo está classificado na subclasse de Poderes do Estado referente ao Direito Constitucional, enquanto a literatura mostra que atos administrativos pertence ao Direito Administrativo. E isso poderá gerar problemas na recuperação da informação, devido a Classificação está incorreta.

Em síntese, a análise e avaliação da representação do domínio em questão pode ser visualizada na tabela 1.

Tabela 1: Representação dos termos na CDU, CDDir, BC e TJ.

TERMOS	CDU	CDDir	BC UFMA	TJ MA
Licitação e Contrato	347.238.8 351.712.2	34.352 341.3527	351.712.2	341.3527
Regime jurídico do Servidor Público	342.98 35.08	341.3313	35.081.7(81)	341.330981
Direitos e garantias fundamentais	342.7	341.27	342.7	341.7

Poderes do Estado	342.5	342.25	342	341.25
Atos administrativos	342.934	341.35	35.0772	341.3501

Fonte: Autor

Ao longo do exercício de análise empreendido, é possível estabelecer algumas recorrências observadas no corpus, que permitem identificar as particularidades entre os dois tipos de Classificação Bibliográfica estudadas. A CDU e CDDir se diferenciam primeiramente na notação, que facilita a localização dos assuntos nas tabelas de classificação. Na CDU a notação é mista, pois utiliza sinais, símbolos, números decimais, sinais gráficos e letras. Já na CDDir a notação é pura, utiliza apenas números arábicos, o ponto serve apenas para facilitar a leitura.

Outra diferença básica entre as duas classificações está na estrutura física, uma vez que a CDDir é composta por uma tabela e um índice compactado em apenas um volume, enquanto a CDU possui dois volumes, sendo um deles o índice sistemático.

A CDU abrange todas as áreas do conhecimento, enquanto que a CDDir é exclusiva da área Jurídica. Por causa disso a CDDir é mais adequada para classificar assuntos desta área.

Na CDU pode-se observar a não existência de uma subclasse própria para o Direito Constitucional, pois ele fica na mesma subclasse do Direito Público, enquanto, que na CDDir existe uma subclasse exclusiva para este termo.

Na CDU foi observado que alguns termos não obtiveram representação dentro do Direito Administrativo ou Constitucional, mas, sim, na Administração Pública.

Na tabela de Classificação da CDDir, todos os termos estudados são condizentes com as representações encontradas na tabela. Os conteúdos programáticos dos concursos estão dentro dos Direitos Administrativos e Constitucionais na literatura, enquanto que nas tabelas, alguns termos na CDU não se enquadram nas subclasses de Direito Administrativo nem constitucional, mas sim na subclasse da Administração Pública. Entretanto não prejudicou a Classificação correta dos documentos na instituição que utiliza a tabela. Na CDDir os termos estão representados dentro das subclasses dos Direitos Administrativo e Constitucional.

A busca da informação é feita de maneira automática, pelas duas instituições, mas como o usuário não está adaptado às classificações, por se tratar de uma atividade específica da Biblioteconomia, existe certa dificuldade na recuperação da informação feita pelo usuário.

CONCLUSÃO

É chegada a etapa de consolidar uma caracterização sobre a importância do uso da Classificação Bibliográfica na organização e recuperação da informação pela biblioteca contemporânea, apresentando uma síntese articulada das constatações a que se chegou ao decorrer das seções anteriores desta monografia, acerca dos Sistemas de Classificação da CDU e CDDir.

A pesquisa propôs descrever e analisar a representação temática do domínio jurídico especificamente dos subdomínios do Direito Administrativo e Constitucional na CDU e CDDir, adotadas pela Biblioteca do TJ-MA e pela Biblioteca Central da UFMA.

Como primeira observação, pode-se dizer que, as Classificações Bibliográficas são essenciais para o trabalho do bibliotecário, seja qual for a categoria de biblioteca que ele trabalhe: pública, especializada, escolar, universitária entre outras. É uma das atividades mais importantes desenvolvida pelo profissional, pois é através dela que a informação é organizada e recuperada, sem ela não haveria como cumprir a principal missão da biblioteca, que é servir de intermediário entre a informação e o leitor, ou seja, a disseminação da informação.

As Classificações Bibliográficas desde suas origens, até os dias atuais, foram criadas, desenvolvidas e adaptadas conforme a necessidade de classificar documentos de forma correta, pois normalmente as classificações abrangem todas as áreas do conhecimento. Mas algumas áreas sentem a necessidade de desenvolver sua própria classificação devido a especificação dos assuntos, como é o caso da Área Jurídica com a CDDir, uma classificação exclusiva da Área Jurídica.

A investigação permitiu um olhar mais aprofundado para a estrutura classificatória dos termos selecionados, e, com isso, possibilitou analisar a maneira na qual profissionais de ambientes distintos tratam documentos semelhantes com instrumentos diferentes, voltados sempre para a necessidade dos usuários e também traz a oportunidade de conhecer e compreender um ambiente que somente os profissionais que trabalham na área, por estarem inseridos nela, têm a possibilidade de conhecer.

Foram apresentadas as origens das Classificações Bibliográficas, seus diferentes tipos foram identificados e caracterizados para uma melhor compreensão.

Buscou-se comparar as representações da Área Jurídica nas tabelas classificatórias CDU e CDDir, em bibliotecas com públicos distintos, para fazer uma análise da forma de classificação, confrontando com a literatura

da área, para contribuir de alguma forma com o processo de organização e recuperação da informação. Diante destes resultados, a estruturação dos conceitos que constituem o *corpus*, da análise desta pesquisa, emerge de dois prismas: Um oriundo da literatura corrente da área jurídica que hierarquiza os conceitos: Licitação e contrato, Regime Jurídico do servidor Público, Direitos e garantias fundamentais, Poder do Estado e Atos Administrativos como termos pertencentes às subclasses do DIREITO ADMINISTRATIVO E DIREITO CONSTITUCIONAL que, por sua vez, constituem a classe do DIREITO PÚBLICO.

O outro ponto de vista advém dos sistemas de classificação CDU e CDDir, como pertencentes às subclasses do DIREITO ADMINISTRATIVO E CONSTITUCIONAL, que fazem parte da classe de DIREITO. Na CDU as subclasses do DIREITO ADMINISTRATIVO E CONSTITUCIONAL são subordinadas à classe do DIREITO PÚBLICO, e os assuntos que lhes são subordinados à elas condizem com a literatura. Na CDDir ambos também pertencem ao DIREITO PÚBLICO, mas, em na subordinação há algumas controvérsias referentes ao DIREITO ADMINISTRATIVO, pois enquanto na literatura o Direito Econômico, Direito Financeiro e Direito Tributário são Subclasses do Direito Público assim como O Direito Administrativo, na CDDir eles estão subordinados ao Direito Administrativo que, por, sua vez, esta subordinado ao Direito Público.

Outra constatação, é que na CDDir os: Direitos Ambiental e Direito Ambiental Internacional estão subordinados ao Direito Administrativo que, por sua vez, está subordinado ao Direito Público, enquanto na literatura, estes direitos pertencem à classe dos Direitos Especiais e não ao Direito Público.

As definições dos termos Licitação e contratos, Regime Jurídico do Servidor Público. Direitos e garantias fundamentais, Poderes do Estado e Atos administrativos, nos dicionários e enciclopédia, ainda que de forma básica, propiciam uma compreensão e análise de sua organização nos Sistemas de Classificação Bibliográfica estudados.

Considera-se, finalmente, que a capacidade dos Sistemas de Classificação em organizar um número significativo de documentos e recuperar a informação precisa em meio de milhares disponíveis, fascina profissionais, pesquisadores e estudiosos que buscam sempre aperfeiçoar essa obra-prima da Biblioteconomia, que é a Classificação Bibliográfica.

REFERÊNCIAS

ALEXANDRINO, Marcelo; PAULO, Vicente. **Direito Administrativo**. 12. ed. Rio de Janeiro: Impetus, 2006.

ANJOS, Liane dos. **Sistemas de Classificação do conhecimento na Filosofia e na Biblioteconomia**: uma visão histórico-conceitual crítica com enfoque nos conceitos de classe, categoria e de faceta. 2008. 290 f. Tese (Doutorado em Ciências da Informação) – Programa de Pós-Graduação em Ciências da Informação, Universidade de São Paulo, São Paulo, 2008.

ARAÚJO, Luiz Alberto David; NUNES JÚNIOR, Vidal Serrano. **Curso de Direito Constitucional**. 15. ed. São Paulo: Verbatim, 2011.

BRAGA, Kátia Soares. Aspectos relevantes para a seleção de metodologia. In: MUELLER, Suzana P.M (org). **Métodos para a pesquisa em Ciência da Informação**. Brasília: Thesaurus, 2007. p. 17-38.

BRASIL. Constituição (1988). Brasília, Senado Federal,Subsecretaria de Edições Técnicas, 2011.

CARVALHO, Dóris de Queiroz. **Classificação Decimal de Direito**. 4.ed. rev. e atual. Brasília, DF: Casa Civil, 2002.

CERVO, Amado L.; BERVIAN, Pedro A.; SILVA, Roberto da. **Metodologia Científica**. 6. ed. São Paulo: Pearson Prentice Hall, 2007.

COTRIM, Gilberto. **Fundamentos da filosofia**: história e grandes temas. 15. ed. São Paulo: Saraiva, 2002.

CRUZ, Anamaria da Costa; MENDES, Maria Tereza Reis; WEITZEL, Simone da Rocha. **A Biblioteca**: o técnico e suas tarefas. Niterói: Intertexto, 2004.

DEWEY, Melvin. **Sistema de classification decimal de Dewey e índice relativo**. 22.ed., edited by Joan S. Mitchel et al. Dublin, Ohio: OCLC Online Computer Library Center, 2003.

DIAS, Eduardo Wense; NAVES, Madalena Martins Lopes. **Análise de Assunto**: teoria e prática. Brasília: Thesaurus, 2007.

DODEBEI, Vera Lúcia Doyle. **Tesouro**: linguagem de representação da memória documentária. Niterói: Intertexto, 2002.

ENCICLOPÉDIA BRITÂNICA [2013].Disponível em:<www.britannica.com>. acesso em: 13 nov. 2013.

FERREIRA, Aurélio Buarque de Holanda. **Mini Aurélio século XXI escolar:** o mimidicionário da língua portuguesa. 4. ed. rev. e aum. Rio de Janeiro: Nova Fronteira, 2001.

FUJITA, Mariângela Spotti Lopes; RUBI, Milena Polsinelli. Um modelo de leitura para a indexação de artigos científicos: princípios de elaboração e uso para a formação de indexadores. **DataGramZero – Revista de Ciência da Informação**, Rio de Janeiro, v. 7, n. 3, p. 1-15, jun. 2006.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2007.

GUINCHAT, Claire; MENO, Michel. A indexação. In---. **Introdução geral às ciências e técnicas da informação e documentação**. 2. ed. Brasília: IBCT, 1994, p. 175-184.

LANGRIDGE, Derek. **Classificação:** abordagem para estudantes de Biblioteconomia. Rio de Janeiro: Interciência, 1977.

MEIRELLES, Hely Lopes. **Direito Administrativo Brasileiro**. 29. ed. São Paulo: Malheiros editores, 2004.

MENDES, Edilze Bonavita Martins. **Visão Panorâmica dos Principais Sistemas de Classificação Bibliográfica**. Campinas: PUCCAMP, 1995.

NAHUZ, Cecília dos santos; FERREIRA; Luzimar Silva. **Manual para normalização de monografias**. 4 .ed. rev. e atual. São Luís: Visionária, 2007.

NAVES, Madalena Martins Lopes; KURAMOTO, Hélio (Org.). **Organização da Informação:** princípios e tendências. Brasília, DF: Briquet de Lemos, 2006.

PIEIDADE, Maria Antonieta Requião. **Introdução à teoria da Classificação**. 2. ed. Rio de Janeiro: Interciência, 1983.

PRADO, Leandro Cadenas. **Licitação e Contratos:** a lei 8.666/93 simplificada. 2.ed. São Paulo: Impetus, 2009.

SANTOS, Washington dos. **Dicionário Jurídico brasileiro**. Belo Horizonte: Del Rey, 2001.

SOUZA, Sebastião de. **CDU:** como entender e utilizar a 2ª Edição Padrão Internacional em Língua portuguesa. 2. ed. Brasília: Thesaurus, 2010.

UDC CONSORTIUM. **CDU**: 2ª Edição-Padrão Internacional em língua Portuguesa. Brasília: IBICT, 2007.

UNIVERSIDADE FEDERAL DO MARANHÃO. Disponível em: <http://www.biblioteca.ufma.br/inf_gerais/missao.php>. acesso em: 21 out. 2013.

PODER JUDICIÁRIO DO ESTADO DO MARANHÃO. Disponível em: <<http://www.tjma.jus.br/tj/publicacoes/sessao/1>>. Acesso em :21 out. 2013.

seja norteada e respaldada por essa filosofia. Assim, pode-se inserir a indexação no contexto administrativo da biblioteca.

Para Rubi (2008), a elaboração de uma política de indexação vai auxiliar a administração da biblioteca no momento da tomada de decisão, quando essa ação desenvolvida representa uma filosofia que reflete os objetivos e os anseios da instituição servida de guia para os bibliotecários, no instante da indexação.

Sendo assim, as bibliotecas universitárias deveriam dedicar uma maior atenção a esta questão, por mais que a instituição disponha de poucos recursos e de um número insuficiente de profissionais bibliotecários pode ocupar-se da política de indexação. Logo, qualquer biblioteca de maior ou de menor porte pode formular sua política utilizando os recursos que dispõe. Desta forma, pretende-se analisar as possíveis causas da falta da política de indexação nas bibliotecas universitárias ludovicenses, apontando os elementos integrantes de tais políticas e indicando a necessidade de sua formalização pelos bibliotecários.

O PAPEL DA INDEXAÇÃO NO SISTEMA DE RECUPERAÇÃO DA INFORMAÇÃO

A indexação é uma das atividades centrais do sistema de recuperação da informação (SRI), uma vez que permite o armazenamento e a recuperação da informação no momento em que o usuário busca um assunto. É uma das técnicas que seleciona ou define termos (palavras ou expressões) que irão representar o conteúdo do documento. A indexação é realizada em dois momentos distintos: na entrada, quando o documento chega à unidade para o tratamento temático da informação, e na saída, quando define a estratégia de busca e recuperação da informação para o usuário.

O processo de indexação é estabelecido com o intuito de facilitar a recuperação da informação, o que proporciona maior precisão e confiabilidade no processo de busca. Todos os sistemas de recuperação da informação são compostos por diferentes etapas que variam de autor para autor, a exemplo de UNIST (1981); Lancaster (2004) que aponta duas etapas; ABNT (1992), que define três etapas; e Robredo (1994), que recomenda quatro etapas.

Mesmo com os autores discordando quanto ao número de etapas do processo de indexação, elas são basicamente as mesmas: Análise Conceitual e Tradução.

A Análise Conceitual tem por objetivo avaliar o assunto tratado no documento. Nessa atividade, a leitura e segmentação do texto são primordiais para identificação e seleção de conceitos, porém o pouco tempo dos profissionais e o grande volume de documentos suscetíveis de tratamento é preocupante. A esse respeito Lancaster (1993, p. 20-21), ressalta que “Ao indexador raramente é dado o luxo de poder ler um documento do começo ao fim [...]”.

Do mesmo modo, Guinchat e Menou (1994) revelam que a indexação é a operação que seleciona termos adequados para representar o conteúdo de um documento, gerando os níveis de indexação que servirão à pesquisa. Esses níveis de indexação podem ser de ordem genérica ou superficial, que identifica o assunto principal do documento tratado, e de ordem exaustiva ou profunda, que pode referir-se à totalidade dos aspectos abordados no documento de forma bastante detalhada.

Nesse âmbito, “O indexador deve extrair todos os termos que julga úteis para descrever completamente todos os conceitos e todos os objetivos tratados no documento capazes de interessar aos usuários de uma unidade documental determinada” é o que recomendam Guinchat e Menou (1994, p.178). A leitura técnica do documento deve adentrar na estrutura, identificando as partes mais relevantes.

Conforme Fujita (1999), os processos de análise e síntese propõem que os textos passem por uma espécie de “desestruturação” para a construção de outro texto: o documentário. Sendo assim, para a análise, os conceitos são identificados e selecionados; na síntese, o texto entra no processo de condensação e na elaboração de um texto documentário, que é um resumo, ou seja, é a produção de um conjunto de frases e enunciados, sintetizando o documento, como palavras-chave e notações de classificação.

É possível conceituar a indexação também como processo de definição de termos que irão representar o conteúdo de documentos em sistemas de informação ou em catálogos de bibliotecas. A indexação é definida como:

[...] A terminologia mais usada para designar o trabalho de instituição da informação quando realizado nos chamados serviços de indexação e resumo. Esses serviços têm por finalidade organizar informações referentes, principalmente, a artigos de periódicos. Essas informações não são, normalmente, organizadas nas bibliotecas. Estas costumam se limitar à instituição dos periódicos considerados em seu todo. Os produtos principais daqueles serviços são os índices/abstracts, que tanto podem estar disponíveis em forma impressa como na forma de bases de dados. (DIAS; NAVES, 2007, p.16).

A etapa de Tradução, segundo Lancaster (1993) objetiva converter a análise conceitual de um documento num determinado conjunto de termos de indexação. Essa mudança ocorre a partir do uso de termos controlados extraídos através do vocabulário controlado. Chaumier (1988, p.41) já previa dizendo que “Os descritores pertencentes a tesouros ou vocabulários controlados constituem-se em linguagens que permitem a tradução para uma linguagem artificial sem ambiguidade denominada linguagem documentária, noções expressas em linguagem natural”. Portanto, devem-se utilizar palavras que funcionam como qualificadores para que o controle dos descritores possibilite a organização e recuperação da informação de forma concisa e eficaz.

Neste contexto, Viera (1988) fala que a indexação é uma técnica de análise de conteúdo que condensa a informação de um documento, através da atribuição de termos, criando uma linguagem intermediária entre o usuário e o documento, que é um dos processos básicos do SRI. Esta noção está vinculada à multiplicidade de fatores intervenientes no ato de indexar realizado pelo bibliotecário.

Segundo Guinchat e Menou (1994), a qualidade da indexação pode ser decidida por vários aspectos, sendo ele exaustivo, quando se verifica todos os temas, e se o conceito tratado no documento foi bem representado; seletivo, quando se constata apenas as informações de interesse do usuário; específica, quando a descrição reduz da forma mais precisa o conteúdo do documento; e uniforme, quando se analisa se os profissionais representam um mesmo documento sobre um mesmo assunto, da mesma forma.

Para Fujita (2006), a indexação necessita de parâmetros que guiem os indexadores no momento da tomada de decisão, diminuindo a subjetividade e incertezas durante esse procedimento. Porém, para o sucesso do processo de indexação é preciso que haja um instrumento regularizador, ou seja, uma política de indexação, uma vez que se constitui em estratégias que permitem o alcance de objetivos de recuperação do sistema de informação. Cabe destacar a importância de estabelecer uma política de indexação para auxiliar o bibliotecário na sua fundamentação na hora de representar o conteúdo temático de um documento. Como se vê, a política de indexação permite ao sistema de informação condições de desenvolver as atividades de representação temática dos documentos de forma racional e consistente, para que o usuário recupere as informações certas, no momento adequado.

POLÍTICA DE INDEXAÇÃO: entre a teoria e a prática

A política de indexação é formada por métodos que devem ser utilizados durante a realização da indexação. Destaca-se como sendo de suma importância em um SRI, agindo como um elo entre o usuário e o documento, interligando-os.

Neste sentido, a política de indexação deve ser constituída de estratégias que permitam o alcance dos objetivos no SRI. A indexação de assuntos constitui-se em uma atividade complexa que deve ser desenvolvida considerando todos os elementos que influenciam direta ou indiretamente no processo: a instituição a que serve, o usuário a que se destina e os tipos de recursos financeiros, materiais e humanos disponíveis. Portanto, no momento de estabelecer uma política de indexação, todos estes condicionantes devem ser levados em consideração. (FUJITA; RUBI, 2006).

De acordo com Carneiro (1985, p.221), uma política de indexação deve servir:

Como um guia para a tomada de decisão, deve levar em conta os seguintes fatores: característica e objetivos da instituição determinantes do tipo de serviço a ser oferecido; identificação dos usuários para o atendimento de suas necessidades e informação e recursos humanos, materiais e financeiros, que delimitam o funcionamento de um sistema de recuperação da informação.

Cesarino (1985) complementa a constatação de Carneiro (1985) explicando que uma política de indexação só pode ser estabelecida depois de observados aspectos como a identificação das características do usuário, da literatura integrada ao sistema, e da qualidade dos recursos humanos e financeiros disponíveis para a criação e a manutenção do sistema. Esses critérios influenciam diretamente na política de indexação.

Para Rubi (2008, p.83) “A política de indexação deve ser compreendida como uma decisão administrativa que reflita os objetivos da biblioteca, identificando condutas teóricas e práticas das equipes envolvidas no tratamento da informação da biblioteca”.

Elementos da política de indexação

Um SRI precisa de uma série de fatores para que os processos nele sejam executados, afetando assim a política de indexação. Para que isso ocorra, são necessários que alguns elementos sejam colocados em prática para uma melhor recuperação da informação. Segundo Guinchat e Menou (1994, p. 180), “[...] o nível de indexação deve ser definido para que estas necessidades possam ser satisfeitas com o menor custo possível”.

Sendo assim, Carneiro (1985) descreve os seguintes elementos que são indispensáveis para uma política de indexação:

a) Cobertura de assunto

Na cobertura de assunto, o profissional da informação, através do estudo do usuário, terá uma ideia dos assuntos cobertos pelo sistema. Irá saber também quais são as necessidades que o sistema precisa para atendê-lo, quais os assuntos centrais e os periféricos, permitindo, desta forma, identificar as áreas que merecem maior destaque para uma cobertura aprofundada, e qual deve ser tratada de modo superficial. Conforme Schardong (2010), a cobertura de assunto deve ser estabelecida junto com a política de Formação e Desenvolvimento de Coleção, que vai determinar o universo temático da comunidade servida pela biblioteca.

b) Seleção e aquisição dos documentos-fonte

Trata-se de um processo primordial ao estabelecimento da seleção e aquisição dos documentos que serão incluídos no sistema. Tendo em vista a extensão de cobertura do sistema, os documentos podem ser incluídos em cada área de interesse, além da qualidade dos documentos para os usuários específicos, o que implicará na determinação de prioridade na aquisição.

Alguns aspectos deverão ser considerados numa política de seleção e aquisição de documentos: aquisição do nível intelectual e técnico dos documentos ao nível dos usuários; domínio, por parte dos usuários, de língua em que os documentos são escritos, limitações de recursos financeiros, o que implicará a determinação de prioridade na aquisição. (CARNEIRO, 1985, p.231).

Dando sequência à afirmação de Carneiro (1985), outro aspecto que deve ser visto diz respeito à qualidade do documento na área de interesse do usuário. E essas áreas podem ser analisadas através de aplicação de questionário, exames das fontes citadas em trabalhos, entre outros onde se possam obter informações de interesse dos usuários.

c) Processo de indexação

No processo de indexação, cada etapa é compreendida por diretrizes que vão influenciar praticamente todo o processo de recuperação da informação, a exemplo da qualidade de indexação, que está diretamente ligada à qualidade de termos atribuídos, à linguagem do vocabulário controlado, entre outros que irão nortear a política de indexação.

- Nível de exaustividade

Lancaster (2004) define o nível de exaustividade como a capacidade de extensão em que os assuntos discutidos no documento são identifi-

cados e determinados na operação de indexação e traduzidos na linguagem do sistema. O nível de exaustividade da indexação é uma decisão política estruturada pela administração do sistema de recuperação, de acordo com a finalidade desta. Em Biblioteca Especializada será maior o nível de exaustividade do que as demais bibliotecas, pois há uma exigência maior.

- Nível de especificidade

Para Lancaster (2004), um importante elemento no processo de indexação é a especificidade, onde se indexa o conceito específico selecionado, ou seja, o assunto é tratado de maneira específica. Assim, o sistema poderá decidir entre uma indexação em maior ou menor profundidade, dependendo da busca que são realizadas. Há uma simetria entre a especificidade e a capacidade de precisão, quanto maior for o grau de especificidade aumenta a proporção de precisão e reduz a revocação.

Foskett (1973 *apud* CARNEIRO, 1985) reforça ainda, dizendo que o nível de especificidade é o resultado de uma decisão administrativa em que a extensão do sistema permite que se analise um dado documento, a fim de especificar o conteúdo temático que se está processando.

- Escolha da linguagem

Segundo Pinto (2001, p.226), linguagem de indexação é “[...] um conjunto de termos estruturados utilizados como tradutores dos elementos indicadores do conteúdo dos documentos, visando à construção de índices para facilitar a recuperação da informação”. Sendo assim, desempenha um papel imprescindível para a recuperação da informação do documento, tanto na estratégia de busca, quando o bibliotecário identifica e descreve os interesses do usuário, quanto na indexação, quando a precisão do bibliotecário descreve o assunto do documento que vai utilizar e recuperar para produzir conhecimento.

Para Lancaster (2004), a escolha da linguagem documentária deve atender às necessidades específicas do usuário, daí porque a instituição deve tomar decisões sobre sua manutenção, atualizando a forma em que o usuário será atendido adequadamente na biblioteca, a fim de facilitar o trabalho do bibliotecário. Para tanto, se faz necessário, no sistema automatizado, optar entre linguagem livre ou linguagem controlada, linguagem pré-coordenada ou pós-coordenada. Na linguagem livre, sua maior vantagem é a rapidez na operação de indexação, o que requer um maior esforço na estratégia de busca. A linguagem controlada é mais lenta, porém o esforço desperdiçado na busca é reduzido.

O referido autor, diz que a linguagem pré-coordenada tem a vantagem de ser mais precisa, o que facilita a estratégia de busca, aumentando o custo da indexação pelo crescimento do arquivo, por repetir os termos na entrada e o uso de referências. A linguagem pós-coordenada é aquela na qual os termos indexados são combinados na hora da recuperação da informação, permitindo uma maior revocação. Apesar de não oferecer uma alta precisão, pode obter associações falsas e revelações incorretas, porém estes fatos podem ser evitados através do aumento de custo tanto no estágio de indexação quanto no de busca. Essas linguagens são adequadas ao sistema automatizado, assim cabe ao bibliotecário escolher a que melhor atenda ao seu usuário.

- Capacidade de revocação e precisão do sistema

Carneiro (1985) ressalta que a capacidade de revocação, o sistema avalia a quantidade de documentos úteis recuperados em uma busca. Já a capacidade de precisão, impede a recuperação de documentos irrelevantes do sistema. Essas variáveis expressam a possibilidade de filtragem do sistema. Enquanto uma é designada a deixar passar o que se solicita, a outra impede o que não é solicitado, portanto elas são inversamente proporcionais, sendo diretamente condicionadas à exaustividade e à especificidade.

Segundo Nunes (2004), a exaustividade, a especificidade, a revocação e a precisão são elementos fundamentais para uma boa política de indexação, assim, tais critérios são essenciais para se determinar a eficiência e eficácia do SRI.

d) Estratégia de busca

Este processo é caracterizado pela busca, de modo delegado ou não, por um especialista. Na busca delegada, a instituição decide quem irá conduzir a pesquisa, se são os bibliotecários responsáveis pelo atendimento, a fim de auxiliar o usuário no sistema. Na busca não delegada, o usuário vai diretamente à base de dado, assim acelerando a sua procura. Outros aspectos podem ser considerados, também, segundo Carneiro (1985 p.236).

Quanto ao esforço que será exigido do usuário na comunicação de sua necessidade de informação, isto é, se será buscado estritamente o assunto na forma em que foi estabelecido pelo usuário, ou se será gasto o tempo que for suficiente para que o técnico de busca se assegure de que entendeu precisamente o pedido. A primeira alternativa minimizará o tempo gasto pelos usuários e pelo técnico de busca no estágio de formulação da busca, mas aumentará o esforço de ambos no processo de seleção dos resultados.

Como afirma Lopes (2002), a rotina de seleção de termos é uma das etapas para elaboração da busca tanto por parte do usuário como do bibliotecário, que precisam seguir diretrizes estabelecidas com relação à estrutura da base de dados. Evidentemente a política de indexação vem para amparar e respaldar a elaboração das estratégias de buscas baseado no manual contendo as instruções de busca a essa base de dados.

Nesse sentido, a tomada de decisão da instituição vai atingir diretamente o estágio de busca realizada pelo usuário, porém o profissional vai conduzir de forma eficiente os mecanismos de busca que beneficiaram o mesmo.

e) Tempo de resposta do sistema

O tempo de resposta é o período utilizado pelo sistema entre o recebimento de uma solicitação de informação e, o fornecimento da resposta tida como aceitável, será determinada em função das necessidades dos usuários a que o sistema se destina que definirá se o tempo gasto foi satisfatório ou não.

Um sistema de alerta corrente exige que as informações sejam fornecidas frequentemente e sejam tão atuais quando possível. Já num sistema de busca retrospectiva, o tempo de resposta pode não ser tão importante para o usuário, uma vez que ele estará mais preocupado com a exaustividade da pesquisa (CARNEIRO, 1985 p.236).

Naturalmente, o tempo de resposta do sistema vai depender da revocação e precisão utilizada pelo usuário. Exigindo uma maior revocação, o sistema vai demorar; se o mesmo estiver interessado em recuperar a informação completa ou se o usuário preferir uma alta precisão, logo exigirá uma resposta mais rápida. A indexação deve presumir através de estudo do usuário, o tempo de resposta satisfatória do sistema, para que se evite o uso de descritores desnecessários.

f) Forma de saída

É o formato em que são apresentados os resultados da busca ao usuário, isto é, a forma de saída do sistema informacional. Essa resposta no sistema pode ser apresentada em forma de número de acesso aos documentos, referências bibliográficas, resumos ou texto completo dos documentos. Assim, para Carneiro (1985, p.237):

O formato tem grande influência sobre a tolerância do usuário quanto à precisão dos resultados. Um usuário que recebe resumos com resultado de sua busca pode tolerar uma precisão mais baixa do que o usuário que recebe os resultados em forma de referências bibliográficas ou apenas números de acesso aos documentos. É mais fácil fazer a seleção a partir dos resumos do que ter que consultar um grande número de documentos para se chegar aos que realmente interessam.

É interessante que se verifique a preferência do usuário no que se refere à apresentação de resultado no momento da busca, pois o usuário tem uma vasta importância no que tange à precisão de resultados.

g) Avaliação do sistema

É um importante fator a ser considerado, pois determinará se o sistema está satisfazendo as necessidades do usuário, se existe alguma falha a fim de detectá-la que devem ser corrigidas.

Segundo Carneiro (1985), a avaliação por mais que se concentre em um determinado subsistema de indexação, este não pode ser avaliado isoladamente, isto porque os vários subsistemas como indexação, busca, linguagem de indexação e interface usuário versus sistema, estão diretamente ligados, portanto qualquer mudança que ocorra em um, conseqüentemente afeta o desempenho do sistema como um todo.

Todos esses fatores citados afetam a qualidade do sistema de indexação e, conseqüentemente, se aplicam à política de indexação que será adotada, no entanto ela deve se atualizar progressivamente. Nunes (2004) complementa dizendo, que há de se admitir diferentes transições dessa política, o que não é admissível é à ausência completa de política documentada formalmente em biblioteca ou serviço de informações de instituições que são razoavelmente dotadas de recursos e que contêm acervos documentais de proporção considerada, ou seja, mesmo uma pequena biblioteca com poucos recursos pode e deve formular sua política de indexação. Dessa forma, a política de indexação vai estabelecer diretrizes para implementação do processo na instituição, seguindo a lista de procedimento a ser adotada pelo sistema de recuperação.

A POLÍTICA DE INDEXAÇÃO NAS BIBLIOTECAS UNIVERSITÁRIAS DE SÃO LUÍS

A biblioteca é um órgão vital na universidade, e um dos seus principais objetivos é sanar as necessidades informacionais de professores, alunos e da comunidade em geral, facilitando assim o acesso, a consulta, e a recuperação de informação para comunidade acadêmica. Nesse sentido, a biblioteca universitária busca desempenhar o seu papel educacional, apoiando as atividades de ensino, pesquisa e extensão da instituição.

A biblioteca universitária, segundo Tarapanoff (1981), deve:

- Preocupar-se com as funções e atividades da universidade a qual pertence;

- Planejar os serviços, relacionando-os aos objetivos de ensino, pesquisa e extensão da universidade;
- Reestruturar suas atividades, em relação às da universidade;
- Restabelecer os seus objetivos de acordo com a política geral da instituição;
- Ter objetivos essencialmente dinâmicos que devem sempre representar as necessidades da universidade a qual pertence.

Desse modo, o propósito da biblioteca universitária é atender com eficiência as necessidades da prática educativa, utilizando meios de mediar o objetivo e a ação através de planejamento, ou seja, oferecendo o acesso à informação (impresa e/ou eletrônica) e promovendo, conseqüentemente, a produção de novos conhecimentos que dará sustento à universidade.

Nas bibliotecas universitárias ludovicenses, a indexação inicia-se pela análise de assunto que é realizada por meio da leitura documentária para atribuir termos ou códigos que serão úteis, posteriormente, na recuperação do documento ou registro. Lancaster (1993) diz, de um modo muito prático, que a 'boa indexação' é aquela que permite a recuperação de termos de uma base de dados, para os quais sejam agilizadas as pesquisas decorrentes dos usuários.

O bibliotecário deve ter conhecimento do conteúdo do assunto tratado e entender sua nomenclatura, embora não precise ser necessariamente especialista no assunto. Lancaster (1993), discorrendo sobre o tema afirma que algumas instituições têm enfrentado dificuldades com indexadores que são muito 'especialistas', pois sua habilidade é traduzir excessivamente e talvez ultrapassar aquilo que o autor afirma. Contudo, a falta de conhecimento do assunto pode conduzir à indexação excessiva. Incapaz de distinguir entre dois termos, o indexador talvez atribua ambos quando apenas um seria necessário ou apenas um seria correto.

Assim Rubi e Fujita (2006) discorrem que é preciso que as bibliotecas compreendam a importância da indexação e do profissional bibliotecário capacitado para estar sintonizando os interesses e as necessidades de informação dos usuários de forma eficiente.

Outro caminho são os sistemas, por mais que se difiram, têm a mesma finalidade no processo de indexação. Os softwares consistem primeiramente na identificação dos termos. Análise do título da obra, o sumário, a introdução, o resumo (quando houver) e a conclusão, é tudo que os bibliotecários declaram ser possível, pois não dá para ler a obra por completo para uma melhor indexação. Segundo Lancaster (2004, p.25), "[...]

todos esses elementos devem ser examinados e avaliados pelo indexador durante a análise que faz do documento". O referido autor não recomenda fazer a indexação utilizando somente o título ou substituindo-o pelo resumo, pois nenhum dos dois são suficientes para expressar o conteúdo do documento, por outro lado a leitura do conjunto de elementos já citados torna a indexação do documento mais completo.

Do mesmo modo, vale salientar a importância da qualidade da indexação, pois vai permitir a recuperação de resposta útil impedindo a resposta inútil na base de dados durante a busca realizada pelo usuário. A qualidade da indexação apenas pode ser aferida com base no resultado de busca alcançada na recuperação, mais essencial, se existir uma avaliação do sistema de recuperação.

Lancaster (2004, p.28), ressalta que não há um conjunto correto de termos de indexação para documento específico. "A mesma publicação pode ser indexada de forma bastante diferente em diferentes centros de informação, e deve ser indexada de modo diferente, se os grupos de usuários estiverem interessados nesse documento por diferentes razões".

De acordo com Lancaster (2004), indexar um documento auxilia na sua localização, pois os termos a ele atribuídos contribuem como pontos de acesso para que os documentos sejam recuperados. É unânime, nas bibliotecas estudadas, a utilização de 6 a 10 descritores por documento, posto que varie de acordo com o tipo de documento com a necessidade informacional dos usuários e as possibilidades das instituições. Os profissionais das instituições estudadas preferem uma indexação exaustiva.

As bibliotecas universitárias ludovicenses realizam algum tipo de avaliação, seja através de aplicação de questionário, de entrevistas ou relatório disponibilizado pelo sistema de busca, com o intuito de saber se o nível de indexação realizado pelos bibliotecários está satisfazendo a ansia informacional dos usuários.

Tais observações levam Rubi e Fujita (2006) a apresentarem soluções para avaliação do SRI através da preparação do usuário para utilização das bases de dados oferecidas pela instituição, por meio de acordo entre os serviços de indexação e serviços de referência para que haja uma melhor adequação da terminologia utilizada, e, por fim, comparando a indexação com seus diferentes serviços, para melhor emprego dos termos através de avaliações.

Nessa perspectiva, a escolha do termo correto na linguagem documental para um descritor diminui a diferença e a incerteza da terminologia utilizada, assim estabelece a eficácia com que se indexa, podendo

representar o conteúdo do documento. Lancaster (2004) considera que discussões relacionadas ao perfil pessoal do bibliotecário têm poder sobre a prática profissional, nesse caso a experiência conta. Porém, o bibliotecário só pode atribuir a um documento termos que constem na lista adotada pela instituição.

Rowley (2002, p. 129) menciona que a estratégia de busca é o “conjunto de decisões tomadas e de procedimentos adotados durante uma busca”. Portanto o bibliotecário deve compreender o sistema que possui em sua biblioteca e saber se os mecanismos oferecidos estão satisfazendo a necessidade do usuário, por meio de questionários e entrevista.

Um sistema de recuperação precisa justificar e comprovar a qualidade de seus serviços. Os estudos tradicionais de desempenho dessas instituições bem como os esforços para seu aperfeiçoamento, têm se orientado para o desenvolvimento das técnicas de processamento da informação: os meios sendo considerados mais importantes que os fins. Entretanto eles são sistemas sociais que devem ser avaliados à luz do processo de comunicação, de um lado, a informação; de outro o usuário; como canal, o sistema de recuperação da informação (CESARINO, 1985, p.166).

O SRI somente expressa sobre a existência ou não, e a localização de documentos relacionados à sua requisição, ou seja, esse classifica os documentos recuperados para futura pesquisa. Dessa forma, por meio da avaliação, a instituição vai adquirir conhecimento do SRI quanto ao assunto central e periférico, além dos níveis de tratamento exigidos, assim como o tipo de resposta, se é exigida maior revocação e precisão ou as duas ao mesmo tempo. Além disso, vai ter conhecimento do nível de exaustividade, de especificidade necessária para linguagem de indexação e o nível de sofisticação desejável no sistema pelo usuário. Portanto isso significa que é fundamental conhecer a clientela, as suas características e identificar seu perfil (CARNEIRO, 1985).

É unânime entre as bibliotecas universitária ludovicense a inexistência da política de indexação estabelecida e formalizada, o que confirma as ratificações de Nunes (2004) sobre a falta de políticas de indexação nas bibliotecas brasileiras. Da mesma forma, Rubi e Fujita (2003) falam que essa deficiência é devido a problemas enfrentados pelos sistemas de bibliotecas do país com relação à normalização de procedimentos de indexação, pois não há manual de indexação, assim como política de indexação.

Nunes (2004) descreve que não há razão porque se deixar de formalizar, estabelecer pequenas regras de procedimentos realizados no dia-a-dia do bibliotecário, como a linguagem de indexação a adotado entre outros aspectos realizados diariamente no processamento técnico.

Observamos que não há uma política de indexação formalizada e estabelecida nas instituições ludovicenses. Os bibliotecários consideram que a política de indexação é uma ferramenta de padronização que garante a qualidade e continuidade dos serviços no processo de indexação. A este propósito, Rubi (2008) salienta que a política de indexação deve ser entendida pela instituição como uma decisão administrativa que represente aos objetivos e que seja reconhecida por condutas teóricas e práticas dos profissionais participantes no tratamento da informação da biblioteca para determinar um padrão de cultura organizacional harmônico com a busca dos usuários.

CONCLUSÃO

A política de indexação tem por objetivo padronizar, aprimorar e auxiliar os instrumentos utilizados na instituição, a fim de assegurar a qualidade informacional e de orientar os profissionais na realização do processamento de indexação e no atendimento da informação desejada ao usuário, sanando por consequente a sua necessidade. Dessa forma deverá estar declarado formalmente em um documento oficial da biblioteca, sendo aprovado pela administração da instituição, para garantir a sua permanência ininterrupta mesmo que haja mudança no quadro de bibliotecários. Nesse sentido, a política de indexação poderá ser atualizada conforme a necessidade da instituição.

Como foram relatadas no decorrer deste trabalho as bibliotecas universitárias ludovicenses não possuem uma política de indexação estabelecida e formalizada, devido ao número reduzido de profissionais e à falta de cursos básicos, especializados, encontros e palestras na área de Biblioteconomia em São Luís, o que aumenta ainda mais as dificuldades desses profissionais em estar se atualizando sobre novas tendências para que se coloque em prática de forma rápida e eficaz para o seu usuário final. Na realidade, mesmo com a falta de atualização, os profissionais têm conhecimento satisfatório na área na qual atuam, assim como são especialistas nos mecanismos de estratégias de busca ofertados pelo sistema.

A indexação nessas instituições é realizada por profissionais capacitados com anos de experiência no processo de indexação. Utilizam aproximadamente dez descritores para representar o conteúdo, preferindo termos que vão do geral ao específico. O nível de indexação pode ser tanto exaustivo como específico, tal preferência depende do tipo de obra e a exigência da comunidade acadêmica. Do mesmo modo a avaliação da indexação é feita através da recuperação da informação no momento da pesquisa realizada pelo usuário e pela aplicação de questionário.

Segundo Rubi (2008, p.92), “o bibliotecário que faz a catalogação do documento deve, portanto, ter consciência da importância da política de indexação e de seus princípios, especialmente a especificidade, a exaustividade, a capacidade de revocação e de precisão do sistema de busca”. Para isso, se faz necessário o emprego de termos satisfatórios para contemplar o conteúdo temático do documento de forma mais completa possível, do qual o profissional deve compreender de maneira generalizada o documento para identificar e selecionar descritores que representarão o mesmo para a recuperação da informação. A escolha do número de descritores pode ser por área, tipo de documento ou localização geográfica, o que fica a critério da instituição.

A política de indexação passa por todo o processo do ciclo documentário, vai deste a coleta informacional até recuperação e disseminação, porém faz-se necessário a criação de um manual que esteja descrita e registrada de maneira clara e objetiva as definições e orientações gerais e específicas aos profissionais, facilitando o desenvolvimento da política de indexação na instituição, pois com o estabelecimento da mesma vai padronizar o processo quanto ao tratamento técnico da informação, condicionando assim um melhor resultado de estratégia de busca na base de dados, permitindo, dessa forma, a sua avaliação e recuperação eficaz. Por outro lado, as discussões sobre a política de indexação ainda encontram-se escassas, portanto faz-se necessário que se estimule novos estudos para que venha fortalecer a literatura na área em estudo.

A esse propósito, o manual de política de indexação deve ser preparado pelas seguintes finalidades: melhorar o atendimento das necessidades informacionais dos usuários da rede de bibliotecas, a dificuldade da tarefa de indexação e a necessidade de padronização de seus procedimentos por parte da instituição; registro dos procedimentos adotados para que, em caso de novo funcionário, a indexação possa continuar sendo realizada da mesma maneira. Esse manual deve fazer parte da documentação oficial da instituição, estar exposto com critérios claros das etapas a serem seguidas para a análise de assuntos, fornecer as regras, diretrizes e procedimentos para o trabalho do bibliotecário e, de preferência, conter os elementos que constituem a política adotada pelo sistema de informação.

Diante do que foi exposto, a política de indexação deve ser inserida no contexto da biblioteca universitária refletindo os anseios dos profissionais envolvidos no tratamento da informação para que dessa forma possa definir um padrão coerente satisfazendo a necessidade dos usuários.

Espera-se com este estudo disponibilizar mecanismos sobre a teoria e a prática da política de indexação, permitindo aos bibliotecários que tenham uma melhor perspectiva da sua realidade, com argumento sobre o ato de indexar e a importância que pode estabelecer ao seu dia-a-dia, evidenciando um impacto direto na transformação de ações e postura frente a novos desafios que lhes são colocados todos os dias. Releva-se a necessidade de desenvolvimento de outros estudos e pesquisas sobre o devido assunto, tendo em vista a complexidade e importância englobada no tema.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12676**: Método para análise de documentos, determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro, 1992.

CHAUMIER, J. Indexação: conceito, etapas e instrumentos. *Revista brasileira de Biblioteconomia e Documentação*. São Paulo, v.21, n.1/2 p.63-79, 1988.

CARNEIRO, Marília Vidigal. Diretrizes para uma Política de Indexação. **Revista da escola de Biblioteconomia da UFMG**, Belo Horizonte, v.14, n.2, p.221-241, set. 1985.

CESARINO, Maria Augusta da Nóbrega; PINTO, Maria Cristina Mello Ferreira. Cabeçalho de assunto como linguagem de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v.7, n.2, p.268-288, set. 1978.

CESARINO, Maria Augusta da Nóbrega; PINTO, Maria Cristina Mello Ferreira. Sistemas de recuperação da informação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, set. 1985.

CONCEIÇÃO, Valdirene Pereira; Pecegueiro, Claudia Maria Pinho de Abreu. Pressupostos teórico-metodológicos da linguagem documentária. **Cad. Pesq.** São Luís, v.13, n.2, p.87-96, jul/dez, 2002.

DIAS, Eduardo Wense; NAVES, Madalena Martins Lopes. **Análise de assunto**: teoria e prática. Brasília: Thesaurus, 2007.

GIL, Antônio Carlos. Como elaborar projetos de pesquisa. 3ed. São Paulo: Atlas, 1991, 159p.

GOMES, **Hagar Espanha**. O indexador face às novas tecnologias de informação. **Transinformação**, Campinas, v.2, n. 1, maio/ago. p.161-17, 1989.

GUIMARÃES, José Augusto Chaves. As políticas de indexação como elemento para a gestão do conhecimento nas organizações. In: VIDOTTI, Silvana Gregório (Coord) **Tecnologia e conteúdo informacionais: abordagens teóricas e práticas**. São Paulo: polis, 2004, p.43-52.

GUIMARÃES, José Augusto Chaves. **A análise documentária no âmbito do tratamento da informação**: elementos históricos conceituais. In: RODRIGUES, Georgete Medleg; LOPES, Ilza Leite. (Org.). *Organização e representação do conhecimento na perspectiva da Ciência da Informação*. Brasília: Thesaurus, 2003, v. 2, p.100-117. Disponível em: <<http://pt.scribd.com/doc/60189087/Analise-documentaria-no-ambito-do-tratamento-da-informacao>>. Acesso em: 11 Set. 2012.

GIL LEIVA, Isidoro; FUJITA, Mariângela Spotti Lopes. **Política de indexação**. São Paulo: Cultura Acadêmica, 2012.

GUINCHAT, Claire; MENO, Michael. A indexação. In: **Introdução geral as técnicas da informação e documentação**. 2.ed. Brasília: IBICT, 1994. p. 175-185.

LANCASTER, Frederic Winfred. **Indexação e Resumos**: teoria e prática. Briquet de Lemos. Brasília, DF: Briquet de Lemos, 1993.

LANCASTER, Frederic Winfred. **Indexação e Resumos**: teoria e prática. Briquet de Lemos. 2.ed. Brasília, DF: Briquet de Lemos, 2004.

LOPES, Ilza Leite. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ci. Inf.**, Brasília, v. 31, n. 1, p. 41-52, jan./abr. 2002.

LOUSADA, Mariana *et al.* Políticas de indexação no âmbito da gestão do conhecimento organizacional. **Inf. & Soc.: Est.**, João Pessoa, v. 21, n. 1, p. 191-202, jan./abr. 2011. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/4136/5607>>. Acesso em: 21 Nov. 2011.

NUNES, Cláudio Omar. Algumas considerações acerca da ausência de políticas de indexação em bibliotecas brasileiras. **Biblos: Revista do Departamento de Biblioteconomia e Historia**, Rio Grande, v. 16, p. 55-61, 2004.

PINTO, Virgínia Bentes. Indexação documentária: uma forma de representação do conhecimento registrado. **Perspect. Cienc. Inf.**, Belo Horizonte, v. 6, n. 2, p. 223 - 234, jul./dez. 2001, p.223-234.

ROBREDO, Jaime. **Documentação de hoje e de amanhã**: uma abordagem informatizada da biblioteconomia e dos sistemas de informação. São Paulo: Global, 1994.

ROWLEY, Jennifer. **A biblioteca eletrônica**. Brasília: Briquet de Lemos/Livros, 2002.

RUBI, Milena Polsinelli; FUJITA, Mariângela Spotti Lopes; BOCCATO, VERA REGINA CASARI. O contexto sociocognitivo do catalogador em bibliotecas universitárias: perspectivas para uma política de tratamento da informação documentária. **Datagramazero**, Rio de Janeiro, v. 10, p. 1-24, 2009. Disponível em: <www.dgz.org.br>. Acesso em: 26 Ago. 2011.

RUBI, Milena Polsinelli. Os princípios da política de indexação na análise de assunto para catalogação: especificidade, exaustividade, revocação e precisão na perspectiva dos catalogadores e usuários. In: **A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias**. São Paulo: Cultura Acadêmica, 2009.

RUBI, Milena Polsinelli. **A política de indexação na perspectiva do conhecimento organizacional**. 2004. 135 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2004.

RUBI, Milena Polsinelli. **Política de indexação para construção de catálogos coletivos em bibliotecas universitárias**. 2008. 166 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2008.

RUBI, Milena Polsinelli. O ensino de procedimentos de política de indexação na perspectiva do conhecimento organizacional: uma proposta de programa para a educação à distância do bibliotecário. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.11, n.1 (jan./abr. 2006), p.48-66.

RUBI, Milena Polsinelli; FUJITA, Mariângela Spotti Lopes. Política de indexação na catalogação de assunto em bibliotecas universitárias: a visão sociocognitiva da atuação profissional com protocolo verbal. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v.7, n. 2, p. 118-150, jan./jun. 2010. Disponível em: <<http://www.sbu.unicamp.br/seer/ojs/viewissue.php?id=20>>. Acesso em: 23 Dez. 2011.

RUBI, Milena Polsinelli; FUJITA, Mariângela Spotti Lopes. Elementos de política de indexação em manuais de indexação de sistemas de informação especializados. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 8, n. 1, p. 66-77, jan./jun. 2003. Disponível em: <<http://www.moodle.ufscar.br/mod/resource/view.php?id=30400>>. Acesso em: 13 Jan. 2012.

SCHARDONG, João Leonel dos Santos. A relevância de elementos para definição de uma política de indexação documental e como eles se apresentam na prática de bibliotecários do sistema de bibliotecas da UFRGS. Disponível em: <http://www.lume.ufrgs.br/bitstream/script=sci_zzarttext&pid=S010019652001000200006>. Acesso em: 13 Dez. 2011.

SEVERINO, Antonio Joaquim. **Metodologia do Trabalho Científico**. 23.ed. revista e atualizada. São Paulo: Cortez, 2007. 304p.

TARAPANOFF, Kira. Objetivos de bibliotecas universitárias. **Revista Latinoamericana de**

Documentación, Brasília, v. 1. n. 1/2, p. 13-17, 1981.

UNISIST. Princípios de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 10, n. 1, p. 83-94, mar. 1981.

VIEIRA, Simone Bastos. **Indexação automática e manual**: revisão de literatura. Ci. Inf., Brasília, 17 (1): 43-57, jan./jun. 1988. Disponível em: <<http://revista.ibict.br/index.php>>.. Acesso em: 11 jun. 2012.

natureza, se o mesmo é um livro, um artigo, seu título, bem como, local e editora de publicação. Em linhas gerais, a Catalogação é a descrição minuciosa de itens de um determinado acervo, de modo a individualizá-lo.

Mey e Silveira (2009, p. 8), complementam, que, ver a Catalogação como uma técnica de elaborar catálogos é algo que ficara no passado, uma vez que é um ideal restritivo e simplista, limitada apenas a listagem de itens. Todavia, ela é capaz de não apenas caracterizar a informação, mas de individualizá-la, torná-la única em meio a um emaranhado de informações, além de possibilitar reuni-las por suas semelhanças. De modo geral, a riqueza da Catalogação está justamente nessa sua capacidade de relacionar os registros, de modo a possibilitar inúmeras alternativas aos usuários.

O fluxo crescente da produção bibliográfica faz a sociedade buscar por uma informação filtrada e de acesso rápido. Essa tendência afeta, sobretudo, aos ambientes informacionais representados pelas bibliotecas, que responsáveis por permitir esse acesso procuram readaptar seus espaços, no intuito de prestar um melhor serviço que atenda às necessidades dos usuários (CASTRO; SANTOS, 2009). A síntese do contexto histórico da Catalogação está no Apêndice A.

O uso intensivo das Tecnologias da Informação e Comunicação (TICs), gera o desenvolvimento de uma gama de padrões de metadados, instrumentos fundamentais nos ambientes de manipulação de dados bibliográficos. Nesse novo cenário, a informação digital passa a fazer parte da vida das pessoas, o universo das bases digitais, enquanto fontes de informação, necessitam serem estruturas, de modo que seu conteúdo possa ser representado, bem como, mecanismos de buscas possam ser elaborados, na perspectiva de garantirem seu acesso. E neste plano que, as estruturas de representação e mecanismos de acesso provocam uma evolução na forma como a catalogação se evidenciava até então.

Ocorreu uma evolução na modelagem e no uso da informação em meio digital, e partindo dessa relação, a catalogação, assume papel de geradora do Tratamento Descritivo da Informação (TDI) e construtora de representação dos recursos informacionais passa a ser associada a usabilidade da informação, além dos próprios metadados, ontologias, modelagem de dados e conteúdo, partindo desde os bancos de dados à própria descoberta e recuperação da informação, e sobretudo o uso de tais recursos (FUSCO, 2011).

O avanço tecnológico, o trabalho técnico dos profissionais da informação, sobretudo do bibliotecário, deve ser cada vez mais preciso e eficiente, de modo a tornar a recuperação do documento com maior ín-

dice de exatidão (PIRES, 2012). A forma de representar essa informação, entenda-se que hoje não materializada apenas no suporte papel, mas em bits e bytes, se apresenta como um grande desafio para os catalogadores.

O avanço das tecnologias, e as progressiva sofisticação e por que não banalização das tecnologias de informação, os bibliotecários e demais profissionais da informação, passaram a trabalhar em si a noção de desadequação, em consequência disso, o subaproveitamento das informações que já existiam nas bases dados. Isso se deve principalmente à rápida mudança que vem ocorrendo nas ferramentas de representação e descrição das informações.

Corrêa (2008), ressalta que, as atualizações foram incorporando os novos suportes informacionais, à medida que estes apareciam, todavia, surgia também a necessidade do estabelecimento de novos padrões conceituais para atender a demanda que vinha sendo gerada pelo avanço frenético das tecnologias de informação casados aos passos curtos dados pela globalização.

Isso desperta nos profissionais da informação, entenda-se aqui Biblioteconomia e Ciência da Informação, a busca e o empenho, no sentido de pensar formas que contemplem e que sejam capazes de representar a informação como ela se apresenta hoje, seja ela em ambiente físico ou virtual, e que essa informação seja recuperada e compreendida (PAES; TABOSA; PINTO, 2011).

Corrêa (2008), destaca que os conceitos estabelecidos pelos FRBR, proporcionará o estabelecimento da recuperação da informação de forma integrada, tornado possível a recuperação de uma obra em todos os itens em que, esta, tiver se manifestado. O que se pretende com os FRBR, é implementar eles já existentes nos catálogos manuais, conhecidas anteriormente por remissivas (simples ou cruzadas), além de estabelecer regras que não deixem brechas para outras interpretações.

O uso das tecnologias no processo de descrição, não mais apenas bibliográfico, mas, que envolve outros tipos de documentos, sendo, necessário assim, entender que a descrição vai além da simples elaboração de um catálogo, mas que perpassa pela organização conceitual da informação por meio das novas ferramentas conceituais. Partindo desse novo cenário, há uma significativa evolução do conceito de Catalogação, como destacam, Fusco e Santos (2009), ela passa do simples processo de descrição da obra em si, para a representação dela, com vistas ao uso, e agora à sua interoperabilidade, se tratando de todo e qualquer tipo de recurso informacional.

Isso implica, em novas provocações à Catalogação, que tem a sua frente não apenas o livro, mas, um número muito maior de materiais a serem descritos, introduzidos a partir das novas tecnologias, e sobretudo, em face da interoperabilidade de bases informacionais heterogêneas, onde se torna, imprescindível compreender as necessidades dos usuários, a preservação destes materiais para posterior uso, e, acima de tudo, o estabelecimento de técnicas de padronizações que permita a recuperação, não apenas em nível local, mas global, uma vez que, vivemos em um mundo com menos fronteiras, onde a cooperação e a troca são palavras de ordem (FUSCO, 2010).

Nesse sentido, são grandes os desafios propostos aos profissionais bibliotecários, na busca por articular os avanços tecnológicos que são inúmeros, e, a cada dia que passa, mais velozes, e sobretudo as possibilidades que tais avanços proporcionam. Destaca-se novamente a necessidade de profissionais mais hábeis e capacitados, que possuam, o mínimo de intimidade que as novas tecnologias de representação da informação. Além do desenvolvimento de novos estudos, na perspectiva de impulsionar o uso de novos métodos de descrição, em face, por exemplo, da informação em meio digital.

METADADOS: tipos e características

O surgimento e desenvolvimento acelerado da *World Wide Web* (WWW), associado aos ganhos em tecnologias, bem como, o aparecimento de uma diversidade de recursos, coloca em evidência a necessidade de se pensarem mecanismos capazes de gerenciar e organizar essa gama de informações disponíveis.

Grácio (2002), ressalta que existem diversos mecanismos de busca na internet, porém nem sempre capazes e eficazes. Um dos principais problemas apontado no processo de busca e recuperação da informação na Web, é justamente a falta de parâmetros, diante dessa infinidade de recursos existentes, o grande volume de informações que é disponibilizado diariamente e sua constante mudança, resultando em um número elevado de documentos, às vezes, não tão relevantes para determinadas pesquisas, ocasionando inclusive a perda de tempo na busca.

Os metadados são elementos intrínsecos e permeiam o processo de catalogação, inclusive no domínio bibliográfico, e fundamentais para os sistemas em ambientes informacionais digitais, pois, “[...] englobam aspectos tecnológicos e [...] representacionais específicos de domínio, promovendo a representação, a individualização, o intercâmbio, a inte-

roperabilidade [...], o acesso e a recuperação de recursos informacionais.” (ALVES; SANTOS, 2013, p. 15).

O termo metadados, é anterior à própria concepção da *Web*, foi criado por Jack E. Myers na década de 1960, nessa época, entendido como “conjunto de dados”. Inicialmente criado com vistas comerciais, com o tempo, passou a ser desenvolvido como teoria para designar a descrição de recursos informacionais.

Alves (2010, p. 47-48), sistematiza e conceitua de forma clara metadados e padrões de metadados:

Metadados são atributos que representam uma entidade (objeto do mundo real) em um sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades; são ainda dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação.

Os **padrões de metadados** são estruturas de descrição constituídas por um conjunto predeterminado de metadados (atributos codificados ou identificadores de uma entidade) metodologicamente construídos e padronizados. O objetivo do padrão de metadados é descrever uma entidade gerando uma representação unívoca e padronizada que possa ser utilizada para recuperação da mesma.

Na figura 1, é possível visualizar como os metadados podem se apresentar sob diferentes aspectos, ambos, no sentido de descrever e individualizar a obra/recurso informacional. Os elementos descritivos presentes nos metadados, também se assemelham com os elementos da própria catalogação, uma vez que, se extraem termos referentes à determinado documento, no sentido de representá-lo e permitir sua recuperação. A ficha catalográfica é um conjunto de metadados, pois, apresenta de forma padronizada e sequenciada elementos descritivos do documento em questão. Assim como a planilha do MARC21, onde apresentam-se os mesmos elementos, verticalmente, também de forma padronizada, e seguindo uma sequência lógica. Nesse sentido, destaca-se como os metadados podem contribuir para o processo de descrição da informação.

Figura 1 – Comparação entre Ficha Catalográfica e MARC21

MEY, Eliane Serrao Alves. **Catalogacao no plural**. Brasília: Briquet de Lemos/Livros, 2009. 217.

025.3
#612CA MEY, Eliane Serrao Alves.
Catalogacao no plural / ELIANE SERRAO ALVES MEY, NAIRA
CHRISTOFOLETTI SILVEIRA. - Brasília: Briquet de
Lemos/Livros, 2009.
217.

1. Catalogacao. 2. Catalogacao descritiva. I. Título

7.5cm

12.5cm

Fechar X

SIGAA | Núcleo de Tecnologia da Informação - UFMA - afonsocunha.ufma.br.sigaa1inst1 v3.7.29-256

Fonte: SIGAA UFMA, 2016

Nessa assertiva, Alves e Santos (2013, p. 43), pontuam que,

[...] os metadados estão presentes no TDI no domínio bibliográfico por meio dos padrões de metadados e outros instrumentos de representação tradicionalmente utilizados no domínio, garantindo a descrição e a representação padronizada dos recursos informacionais.

Tais padrões tem por finalidade garantir, não apenas a descrição, mas, também a representação padronizada dos recursos informacionais, independente de seu suporte, na perspectiva de tornar fácil o processo de organização, busca e recuperação da informação.

Alves e Santos (2013), fecham o quadro de princípios, destacando outros requisitos essenciais para o desenvolvimento de metadados, estes por sua vez, mais abrangentes e específicos, os princípios são os seguintes: modularidade; extensibilidade; refinamento; multilinguismo; perfil de aplicação; sintaxe e semântica; associação entre modelos; identificação e designação de elementos de metadados; registro de metadados; integralidade na descrição; elementos obrigatórios e opcionais; metadados objetivos e subjetivos; geração automática de metadados. Todos esses princípios e padrões, devem estar em consonância, no sentido de proporcionar uma representação adequada e garantir a interoperabilidade entre os sistemas.

O Dublin Core, se apresenta como um padrão de metadados com propósitos gerais. É composto de um sistema de metadados simples, e ao mesmo tempo bem eficaz, com o intuito de descrever uma infinidade

de recursos, nos mais distintos ambientes informacionais na Web (ALVES; SANTOS, 2013).

O conjunto de metadados descritos pelo Dublin Core, é formado por 15 elementos, os quais podem ser descritos como o mais baixo denominador comum para a descrição de recursos informacionais, de certa forma, pode ser comparado à uma ficha catalográfica, todavia, o DC não visa substituir modelos já consagrados como o AACR2, todavia, visa conceder um conjunto de informações básicas de cunho descritivo, que podem ser usados por catalogadores no processo de descrição (SOUZA; VENDRUSCULO; MELO, 2009).

Segundo Alves e Santos (2013), o esquema de metadados do padrão Dublin Core, é constituído de dois níveis simples e qualificado.

a) nível simples – composto por 15 elementos de descrição;

b) nível qualificado – apresenta além dos quinze elementos do nível simples, inclui mais sete elementos e duas classes de classificadores: elementos de refinamento e elementos de codificação.

Segundo Lourenço (2005, p. 63-64), o *Dublin Core* possui atualmente 15 elementos descritivos que são:

1. *Title* – título do objeto
2. *Creator* – responsáveis pelo conteúdo intelectual do objeto
3. *Subject* – tópico relacionado ao objeto descrito
4. *Description* – contém uma descrição textual do objeto
5. *Publisher* – agente responsável por tornar o objeto disponível
6. *Contributor* – outros “autores” do conteúdo intelectual do objeto
7. *Date* – data de publicação
8. *Type* – tipo do objeto
9. *Format* – formato de dado do objeto
10. *Identifier* – identifica o recurso de forma única
11. *Source* – objetos dos quais o objeto descrito é derivado
12. *Language* – idioma relativo ao conteúdo intelectual do objeto
13. *Relation* – indica um tipo de relacionamento com outros objetos
14. *Coverage* – localização espacial e duração temporal do objeto
15. *Rights* – contém referencia ou direitos de propriedade

As principais características do Dublin Core, é a simplicidade na descrição dos recursos, o entendimento semântico universalizado que o padrão proporciona, seu escopo que tem alcance internacional, e sua capacidade de extensibilidade, permitindo sua adaptação às mais distintas necessidades de descrição.

MARC21 e Metadados

Segundo Modesto (2008), o MARC consiste num sistema onde o computador é capaz de interpretar os dados de um determinado registro bibliográfico, e assim, permitir o intercâmbio de registros bibliográficos. Segundo Silva e Baptista (2013), um determinado registro bibliográfico quando utilizado o formato MARC, é composto por três elementos principais: o líder, o diretório e os campos variáveis. A divisão básica dos campos do MARC 21 é:

- 0XX Informações de controle, números e códigos
- 1XX Entrada principal
- 2XX Título, edição, impressão (em geral, o título, a indicação de responsabilidade, a edição e as informações da publicação, distribuição etc.)
- 3XX Descrição física
- 4XX Designação de série
- 5XX Notas
- 6XX Entradas adicionais de assunto
- 7XX Entradas adicionais de outros assuntos ou séries
- 8XX Entrada adicional de série (outras formas de autoridades)
- 9XX Destinado para uso de decisões locais.

Modesto (2008) destaca que, todo e qualquer tipo de documento pode ser descrito, nesse sentido, os seguintes tipos de materiais contemplados pelo MARC:

- a) BK (*Books*) – livros, teses, dissertações, etc;
- b) CF (*Computer Files*) – arquivos de computador;
- c) MP (*Maps*) – mapas, globos, etc;
- d) MU (*Music*) – discos, fitas, cds;
- e) SE (*Serials*) – periódicos;
- f) VM (*Visual Materials*) – fotos, posters, slides, vídeos etc;
- g) MX (*Mixed Materials*) – kits que misturem mídias, material de arquivo.

Apresentando uma estrutura de metadados rígida, o MARC 21, passa por constantes atualizações, na perspectiva de tornar o processo de descrição mais detalhado. Por ser um padrão rígido, foi necessário que se criasse um padrão mais flexível, de modo a contemplar recursos em ambiente *Web*, o MARC XML (ALVES, 2010). O MARC XML objetiva resolver esse problema no que diz respeito à flexibilidade, extensibilidade e

modularidade dos recursos informacionais na *Web*, sem que perca a especificidade de domínio, assegurando que a interoperabilidade entre os sistemas não seja afetada (ALVES; SANTOS, 2013).

Lourenço (2005), ressalta que, outra instituição que tem se sobressaído no estudo de padrões de metadados é a *World Wide Web Consortium* (W3C) que vem desprendendo esforços no desenvolvimento do padrão *Resource Description Framework* (RDF), padrão baseado na linguagem de marcação XML, também um produto da W3C que tem como principal foco de pesquisa o estudo e aperfeiçoamento da semântica na web.

MODELOS CONCEITUAIS DE REPRESENTAÇÃO DA INFORMAÇÃO

O desenvolvimento de sistemas automatizados com vistas ao processamento de dados bibliográficos, possibilitou o crescimento em larga escala das bases de dados, e o ideal da catalogação compartilhada emergiu desde então. Outro fator que merece destaque, é o demasiado crescimento da produção editorial, com a ascensão das editoras eletrônicas, e das redes que permitiam o acesso às informações agora em rede.

Necessitava-se que fosse otimizado o tratamento dessa informação, fruto então da cooperação entre as bibliotecas e centros de informação, uma vez que, vinham-se sendo criados bancos de dados remotos, que armazenavam, desde as bases de dados referenciais, até os próprios repositórios documentais. Sendo assim, era necessário que seus metadados fossem padronizados, de modo a permitir e garantir a recuperação da informação, e sobretudo, a interoperabilidade dos sistemas de informação, além do uso e reuso das informações e dos documentos que vinham sendo disseminados (BEZERRA; SOUZA, 2011).

Requisitos Funcionais para Registros Bibliográficos: FRBR

Diante de novos contextos e novos ambientes onde a informação passara a se manifestar, implicaram em uma intensificação nos custos de catalogação, e sobretudo, diante da necessidade de que se economizasse nesse processo, em face do vasto crescimento das publicações e o surgimento vertiginoso de novos formatos, novos suportes, além das mais distintas formas de acesso, implicaram em estudos que gerassem métodos capazes de lidar com essa nova realidade (MORENO; ARELLANO, 2005).

Em face da nova realidade informacional e da necessidade cada vez maior de uma catalogação cooperativa, a IFLA criou um grupo para estudar as práticas das normas de Catalogação, tanto que no período de 1992 a 1995, desenvolveu um modelo conceitual de natureza entidade-relacionamento, tendo sua publicação realizada em 1998. Os Requisitos

Funcionais para Registros Bibliográficos, são, por sua vez, é independente de qualquer código de catalogação ou implementação.

Portanto, o modelo conceitual FRBR, não se trata apenas de um código de catalogação, quanto um formato destinado aos registros bibliográficos, mas sim um modelo entidade-relacionamento, cuja proposta é apresentar requisitos essenciais para um determinado registro bibliográfico e os relacionamentos que este pode estabelecer como suas entidades, seja por meio da relação obra-expressão, expressão-manifestação, manifestação-item (PACHECO, 2014).

A IFLA (1998), também destaca dois objetivos fundamentais dos FRBR, o primeiro deles é fornecer um quadro estruturado e claro, de modo a proporcionar o relacionamento de dados registrados e registros bibliográficos às principais necessidades dos usuários destes; e o segundo é indicar um nível básico de funções para os registros criados pelos órgãos bibliográficos internacionais. Sendo assim, os FRBR foram pensados para atender as mais distintas necessidades dos usuários, de modo que estes, se modelam às tarefas básicas, realizadas pelos usuários, na busca por informação.

Tais buscas por informação se materializam nas *user tasks*, que podem ser chamadas de tarefas genéricas ou básicas. A IFLA (1998) também descreve de modo bem claro cada uma dessas tarefas. As atividades genéricas são: encontrar, identificar, selecionar e obter.

a) **encontrar** – corresponde aos critérios estipulados para a busca realizada pelo usuário, as entidades que dizem respeito aos critérios indicados na mesma, ou seja,

b) **identificar** – consiste em confirmar que a entidade descrita, diz respeito à entidade procurada, ou, distingui-la entre duas ou mais entidades de mesmas características;

c) **selecionar** – nada mais é que selecionar uma entidade que atenda aos requisitos do usuário, tais requisitos podem se relacionar com o conteúdo, formato, entre outros, e também, pode levar a rejeição da mesma por parte do usuário, uma vez que não atenda às suas necessidades;

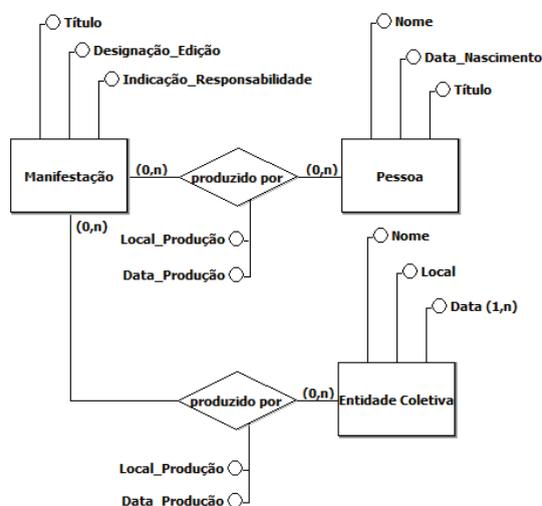
d) **obter** – refere-se ao acesso à entidade descrita, nada mais é que adquirir por meio de empréstimo, compra, entre outras formas, bem como o acesso por meio eletrônico, fazendo uso de internet ou computador.

Sendo assim, os FRBR têm por finalidade beneficiar as tarefas dos usuários em sistemas automatizados. Conforme Fusco (2010), ele é considerado modelo conceitual, uma vez que, representa e descreve teoricamente o universo bibliográfico, servindo assim, de base para a implemen-

tação de diferentes serviços ou bases de dados bibliográficas. Silva (2014, p. 32), continua dizendo que, os FRBR, proporcionam uma recuperação mais eficiente dos itens documentais, “[...] pois ligam todos os materiais disponíveis ao termo da busca em uma única interface na base de dados.” Dessa forma, o usuário economiza tempo, bem como, tem à sua disposição uma gama diversificada de documentos, nos mais distintos suportes.

Os FRBR têm em sua base o Modelo Entidade-Relacionamento (ER), esse modelo tem por finalidade reorganizar os registros bibliográficos, na perspectiva de reordenar os elementos informacionais, a partir, da análise de suas entidades, atributos e possíveis relacionamentos (FUSCO, 2010; MORENO 2006), conforme figura 2:

Figura 2 – Diagrama Entidade-Relacionamento



Fonte: Fusco (2010)

O Modelo E-R congrega uma técnica particular de diagramação, esta que por sua vez, não é utilizada nos FRBR, essa técnica é conhecida por Diagrama Entidade-Relacionamento, porém, é bastante útil na explicação do modelo. Neste mote, o modelo conceitual proposto para descrever estruturas bibliográficas utilizando conceitos dos FRBR, é embasado nas técnicas do Diagrama Entidade-Relacionamento, ora citado, e, é capaz de retratar graficamente o que pode ser um banco de dados baseado no modelo FRBR (FUSCO, 2010).

Os FRBR contam ao todo com dez entidades, estas por sua vez, são os objetos de interesse dos usuários mediante suas necessidades infor-

macionais. Elas são peça-chave, uma vez que, proporcionam um novo tipo de abordagem para os registros do conhecimento, sendo assim de interesse fundamental dos usuários (BEZERRA, 2013).

Essas dez entidades ou classes gerais dividem-se em três grupos: **Grupo 1:** Produto de trabalho intelectual ou artístico; **Grupo 2:** representa os responsáveis pelo conteúdo intelectual ou artístico; e, **Grupo 3:** Conjunto adicional de entidades que servem como assuntos de Obras

As entidades do **Grupo 1**, são as descritas nos registros bibliográficos, diz respeito aos diferentes aspectos dos interesses dos usuários dos materiais produzidos tanto de natureza intelectual, quanto artística, conforme figura 3. São elas:

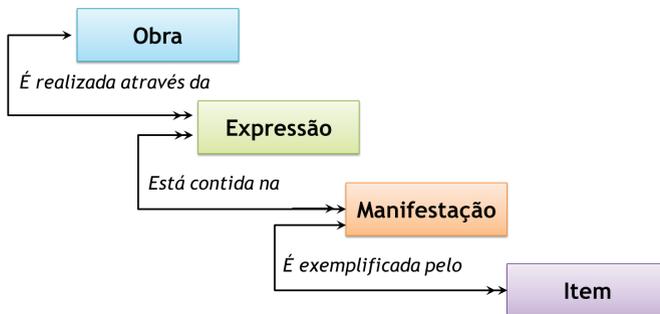
a) Obra – criação intelectual ou artística distinta;

b) Expressão - a realização intelectual ou artística de uma obra, esta ao ser realizada, desprezando-se aspectos de mudança de sua forma física, ou seja, refletem o conteúdo intelectual ou artístico;

c) Manifestação – é a materialização de uma expressão de uma obra, nada mais é que seu suporte físico, podendo se apresentar sob a forma de um livro, jornal, filmes, dentre outras formas;

d) Item – um único exemplar de uma determinada manifestação.

Figura 3 – Entidades do Grupo 1 e seus relacionamentos

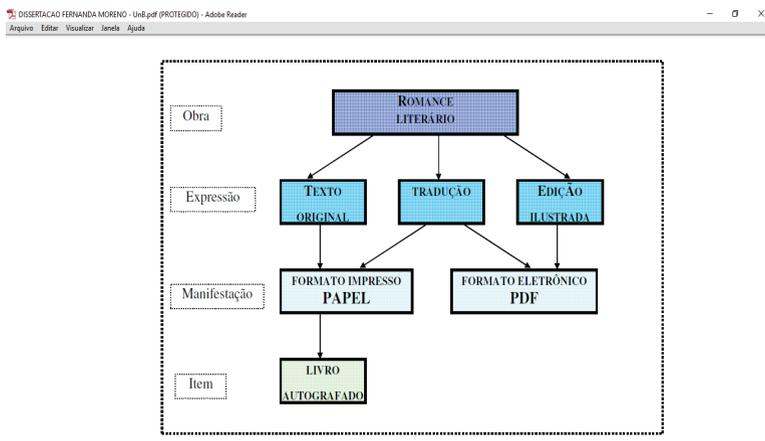


Fonte: Fusco (2010) e Silva (2014).

Uma obra pode ser realizada mediante uma ou mais expressões, todavia, uma expressão é realização de apenas uma obra. Uma expressão está contida em uma ou mais manifestações, bem como, uma manifestação pode conter mais de uma expressão. Uma manifestação pode ser exemplificada por um ou mais itens, entretanto, um item exemplifica apenas uma manifestação (SILVA, 2014). Fusco (2010), ressalta que, Obra e Expressão são responsáveis por refletirem o conteúdo intelectual

ou artístico, e que Manifestação e Item, enquanto entidades concretas, incumbem-se de refletir a forma física desses. Na figura 4, é apresentado um exemplo de descrição mediante os FRBR:

Figura 4 - Exemplos Grupo 1



Fonte: Moreno (2006)

Abaixo são apresentados outros dois exemplos, o primeiro – na figura 5 – apresentação a descrição do Jornal O Estado do Maranhão. A obra no caso é o Jornal O Estado do Maranhão, expresso na edição 23, e que pode se manifestar tanto no formato impresso ou no formato eletrônico, um item dessa manifestação pode ser encontrado numa banca de jornal, conforme exemplo:

Figura 5 – Exemplo do Jornal O Estado do Maranhão

Obra: Jornal O Estado do Maranhão

Expressão: Edição 23

Manifestação 1: formato impresso

Manifestação 2: formato online

Item: Impresso adquirido na banca Deodoro.

Fonte: Elaborado pelo autor

Figura 6 – Exemplo do Filme do Harry Potter

Obra: Filme Harry Potter e a Pedra Filosofica

Expressão 1: a versão com idioma original (Inglês)

Expressão 2: a versão dublada em Língua Portuguesa

Obra 2: Peça de Teatro Harry Potter e a Pedra Filosofica.

Fonte: Elaborado pelo autor

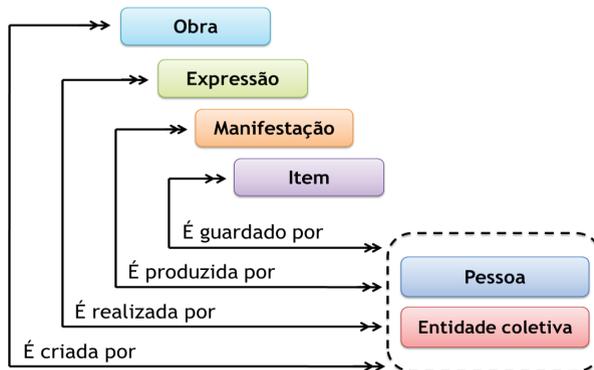
No segundo exemplo, na figura 6, é apresentada a descrição do Filme Harry Potter e a Pedra Filosofal, que pode ser expressa na versão original em inglês ou dublada em língua portuguesa, e pode ter uma segunda obra, a Peça Teatral.

Grupo 2: representa os responsáveis pelo conteúdo intelectual ou artístico, pela produção física e disseminação ou pela guarda das entidades do primeiro grupo. São elas, conforme figura 7:

a) Pessoa – um indivíduo ou pessoa física responsável pela obra. Esta entidade pode ser identificada como sendo o autor, editor; compositor; artista, diretor; interprete; tradutor;

b) Entidade coletiva ou organizações – uma organização ou grupo de indivíduos responsáveis pela obra.

Figura 7 – Entidades do Grupo 2 e seus relacionamentos



Fonte: Adaptado de Fusco (2010) e Silva (2014).

Como evidenciado na figura 7, que uma obra pode ser criada por uma ou mais pessoas ou entidades coletivas, assim como uma ou mais pessoas ou entidades coletivas podem ser responsáveis por uma ou mais obras. Uma expressão pode ser realizada por uma ou mais pessoas ou entidades coletivas, assim como, estas podem realizar uma ou mais expressões.

A manifestação pode ser produzida por uma ou mais pessoas ou entidades coletivas, assim como estas podem se incumbir de produzir uma ou mais manifestações. E finalmente, um item pode ser de propriedade de uma ou mais pessoas ou entidades coletivas, assim como, uma ou mais pessoas ou entidades coletivas podem possuir um ou mais itens (SILVA, 2014; IFLA, 2007).

Grupo 3: Conjunto adicional de entidades que servem como assuntos de Obras: conceito, objeto, evento e lugar, conforme evidenciado na figura 8. A IFLA (2009), conceitua cada uma das entidades da seguinte forma:

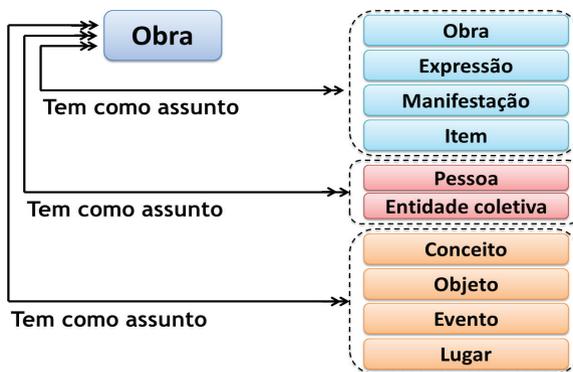
a) Conceito – compreende diferentes abstrações pertinentes a determinada Obra, podendo ser as áreas do conhecimento, as próprias disciplinas, processos, técnicas, práticas, entre outros;

b) Objeto – compreende uma série de objetos, tanto animados quando inanimados, podendo ser elementos da natureza, objetos fixos, móveis, entre outros;

c) Evento – contempla as ações ou fatos que por ventura possam ser objetos de uma Obra, entre eles, acontecimentos históricos, épocas, períodos do tempo, entre outros.

d) Lugar – diz respeito à localidade, posição geográfica, características geográficas e jurisdições geopolíticas.

Figura 8 – Entidades do Grupo 3 e seus relacionamentos



Fonte: Fonte: Adaptado de Fusco (2010) e Silva (2014).

Na figura 8, são apresentadas as entidades dos três grupos (estas apresentadas em cores diferentes, Grupo 1: azul; Grupo 2: vermelho; Grupo 3: laranja), enaltecem-se as relações de assunto, e também é possível perceber que uma obra pode ter como assunto todas as entidades dos três grupos, sendo possível que ocorra de uma ou mais delas na mesma obra.

As entidades do Grupo 3, podem se relacionar, por exemplo, com as tabelas auxiliares da CDU. A entidade “objeto” se assemelha à auxiliar de

“forma”, a entidade “evento” à auxiliar “tempo”, a entidade “lugar” à auxiliar de “lugar”.

As entidades dos FRBR possuem um conjunto de atributos, estes por sua vez, são os meios pelos quais os usuários pensam suas buscas, recuperam e interpretam as respostas na busca por informação. Os atributos permitem definir as entidades, bem como distinguir uma Obra de outra, ou uma Obra e uma Expressão, entre duas Expressões, além de diferenças no próprio conteúdo intelectual ou artístico (IFLA, 1998).

Silva (2014), apresenta de modo sintético os atributos das entidades dos FRBR. No quadro 1, são apresentados os atributos do Grupo 1:

Quadro 1 – Atributos das entidades do Grupo 1

Obra	Expressão	Manifestação	Item
Título da obra	Título da expressão	Título da Manifestação	Identificador do item
Forma da obra	Forma da expressão	Indicação de responsabilidade	Impressão digital
Data da obra	Data da expressão	Local de publicação/ distribuição	Proveniência do item
Público a que se destina	Língua da expressão	Publicador/ distribuidor	Condição do item
Contexto da obra	Contexto para a expressão	Data de publicação/ distribuição	Restrições de acesso ao item

Fonte: Adaptado da obra de Silva (2014)

Os atributos das entidades do FRBR, são de grande importância para os registros bibliográficos, uma vez que, evidenciam, não apenas os entes descritivos, como também, os pontos de acesso, itens de organização, bem como notas. No quadro 2, são apresentados os atributos do Grupo 2:

Quadro 2 – Atributos das entidades do Grupo 2

Pessoa	Entidade Coletiva
Nome da pessoa	Nome da entidade coletiva
Datas da pessoa	Número associado à entidade coletiva
Título da pessoa	Local associado à entidade coletiva
	Data associada à entidade coletiva

Fonte: Adaptado da obra de Silva (2014)

E no quadro 3, são apresentados os atributos das entidades do Grupo 3:

Quadro 3 – Atributos das entidades do Grupo 3

Conceito	Objeto	Evento	Lugar
Termo para o conceito	Termo para o objeto	Termo para o Evento	Termo para o lugar

Fonte: Adaptado da obra de Silva (2014)

Tais atributos, podem ser definidos por um ou mais termos, bem como por uma ou mais formas. As unidades de informação por sua vez, também desmembram tais atributos, uma vez que definem termos para os cabeçalhos, e os demais termos são considerados variantes do mesmo (SILVA, 2014; IFLA, 1998).

Requisitos Funcionais para Dados de Autoridade: FRAD

Bezerra (2013), ressalta que o modelo conceitual FRAD, tem como finalidade controlar os dados de autoridade, sendo de grande importância, sobretudo, para os catálogos. Estes por sua vez, reúnem informações acerca de determinada informação, sob controle de uma instituição, e, é por meio deles que essas informações são recuperadas.

Segundo a IFLA (2009), os dados de autoridade nada mais são do que pontos de acesso controlados acerca de determinada pessoa, família, entidade coletiva, edições de um mesmo título. Esses pontos apresentam tanto formas autorizadas, quanto variadas de nomes, definidos pelos catalogadores na perspectiva de identifica-los, sendo de grande importância para a plena funcionalidade dos catálogos.

Sintetizando, o objetivo central do modelo, nada mais é que, disponibilizar uma análise dos requisitos funcionais para dados de autoridade, tanto na perspectiva de controlar tais dados, quanto de proporcionar um intercâmbio internacional de dados de autoridade. Sendo assim, visa-se fornecer, um quadro claro e estruturado relacionando os dados coletados pelos catalogadores face às necessidades de que os utilizam (IFLA, 2009; BEZERRA, 2013).

Como fruto dos FRBR, o FRAD também faz uso da metodologia presente no modelo entidade-relacionamento (E-R), falado anteriormente. Sendo assim, apresenta suas entidades, bem como seus atributos e seus relacionamentos. Sua base está nas entidades do universo bibliográfico presente no modelo FRBR.

Figura 9 – Base do modelo conceitual FRAD

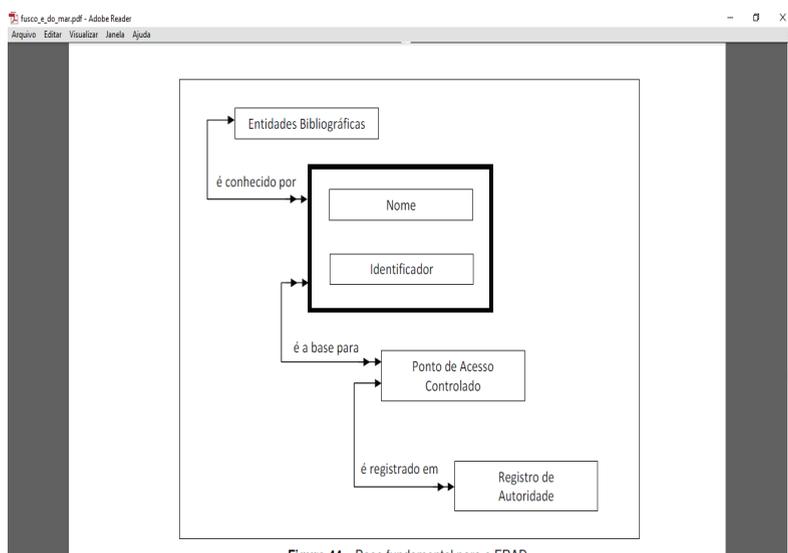


Figura 10 – Base fundamental para o FRAD

Fonte: Fusco (2010).

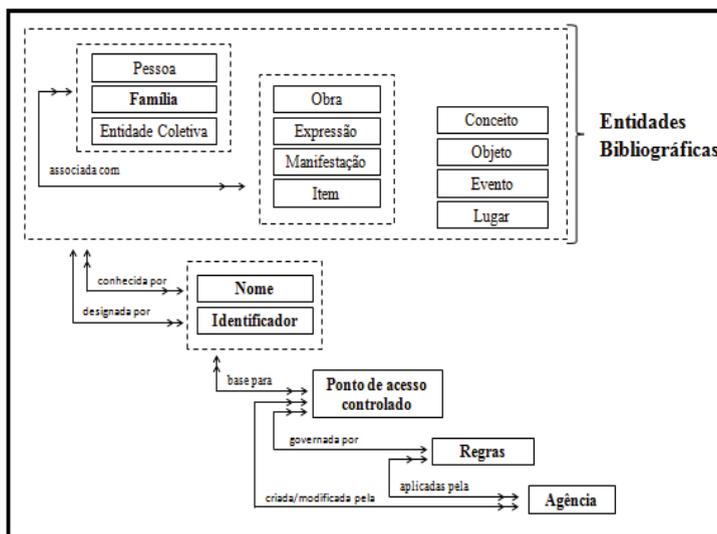
Esse controle de autoridade também é expresso no próprio AACR2, na definição dos pontos de acesso, onde, são definidas as entidades bibliográficas, o autor da obra, que por sua vez pode ser autor pessoal ou autor entidade, estabelecendo assim, o ponto de acesso controlado, expresso pelos FRAD.

a) Entidade Nome - O FRAD não inclui "Título" como atributo da Obra, Expressão ou Manifestação, sendo assim, trata o "Nome" ou "Título" da Obra, Expressão ou Manifestação como uma entidade.

b) Entidade Identificador - Definida pela IFLA (2009), como sendo um número, código, palavra, frase, logo, entre outros, cuja associação é feita à uma entidade, na perspectiva de diferenciar esta entidade das demais, no ambiente ao qual o identificador faz parte. Um exemplo de identificar é um número de código de barras, bem como, qualquer entidade presente nos FRBR.

Na figura 10, são é apresentado o modelo conceitual e suas entidades:

Figura 10 – Modelo conceitual para dados de autoridade



Fonte: Bezerra (2013)

Tanto FRBR, quanto FRAD, tem por finalidade a padronização dos elementos descritivos no processo de Catalogação, ambos contribuem significativamente para a expansão da semântica, sendo pontual na elaboração de catálogos voltados para o ambiente virtual.

Requisitos Funcionais para Dados de Autoridade Assunto (FRSAD)

Segundo a IFLA (2010), o FRSAD foi desenvolvido, uma vez que, os FRBR não contemplavam plenamente os dados de autoridade – tendo em vista a criação de um outro modelo conceitual que o fizesse, no caso o FRAD – criando-se então o modelo FRSAD, buscando abraçar todos os tipos de dados de autoridade. Na busca por alcançar seus principais objetivos, o Grupo de Trabalho FRSAD foi composto por dois subgrupos:

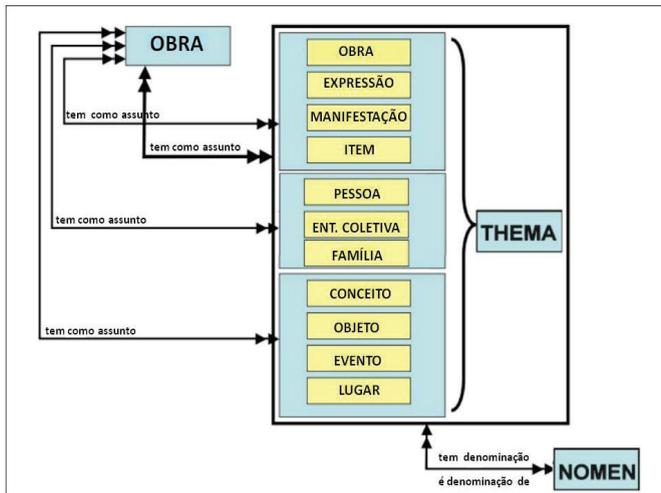
a) *User Tasks Sub-Group* – responsável por estudar as tarefas do usuário de dados de autoridade e assunto;

b) *Subject Entities Sub-Group* – incumbido de estudar as entidades do Grupo 3 do FRBR, como ressalta Bezerra (2013, p. 107), tinha como objetivo “[...] definir as entidades que podem servir como assunto de uma obra; possíveis sub-entidades do conjunto do Grupo 3 do FRBR; e entidades adicionais relacionadas com o conjunto de entidades do Grupo 3 do FRBR.”

As tarefas de usuário do FRSAD foram redefinidas, além das tarefas básicas ora apresentadas nos FRBR - encontrar, identificar e selecionar – neste modelo, insere-se uma nova tarefa, “explorar”; esta, realizada pelo usuário em sua busca (MELO; BRASCHER, 2014).

Mantendo a linha E-R, os FRSAD também é estruturado em entidades, atributos e relacionamentos. Conforme figura 11:

Figura 11 – Relação entre os modelos FRSAD e FRBR



Fonte: Bezerra (2013); IFLA (2010)

Como pode-se observar, o FRSAD apresenta a relação entre Obra e todas as entidades presentes nos três grupos dos FRBR, acrescida do grupo de entidades do FRAD. A diferença é que nos FRSAD, apresentam-se duas novas entidades:

a) *Thema* – que corresponde à qualquer entidade que pode ser utilizada como assunto de uma determinada obra;

b) *Nomen* – entidade definida como sendo qualquer signo ou sequência de signos – podendo ser caracteres alfanuméricos, símbolos, entre outros – pelo qual um *thema* é “conhecido por, referenciado a, ou tratado como” (IFLA, 2010). Silva (2014), apresenta de forma condensada em sua obra os principais atributos das duas entidades do FRSAD, bem como os conceitua, conforme pode ser visualizado no quadro 4:

Quadro 4 – Atributos das Entidades do FRSAD

ENTIDADES	ATRIBUTOS	CONCEITO
THEMA	Tipo de Thema	Qualquer entidade bibliográfica pertencente à esta categoria, uma vez que um thema é de propriedade de uma organização ou sistema particular.
	Nota de alcance	Texto que descreve/define um thema, bem como especifica seu alcance.
NOMEN	Tipos de nomen	Aspectos importantes deste atributo: a) identificador – nome atribuído à uma entidade dentro de um determinado domínio; b) nome controlado – nome atribuído mediante controle de autoridade ou uso em vocabulário controlado como ponto de acesso.
	Esquema	Onde se estabelece o nomen, por exemplo, listas de cabeçalho de assunto, tesouros, etc.
	Fonte de referência	Fonte onde se encontra o nomen
	Representação	Tipo de dados os quais se expressa o nomen
	Idioma do nomen	Idioma em que se apresenta
	Alfabeto	Alfabeto no qual está escrito
	Tradução	Sistema ou norma de tradução
	Forma	Informação adicional que contribua para sua interpretação
	Tempo de validade	Período de tempo para que o nomen seja valido ou utilizada em vocabulários controlados
	Público	Forma predileta para determinado grupo de usuários
	Status	Seu status em um sistema de autoridade de assunto

Fonte: adaptado de Silva (2014)

Melo e Brascher (2014), sintetizam os relacionamentos presentes no modelo conceitual FRSAD. Este modelo estabelece dois conjuntos básicos de relacionamento, como pode ser visto na figura 9: a) entre diferentes tipos de entidades – *obra-to-thema*; *thema-to-nomen*, e, b) entre entidades de mesmo tipo – *thema-to-thema*; *nomen-to-nomen*. Conforme figura 12:

Figura 12 – Relacionamentos do Modelo conceitual FRSAD



Fonte: Melo e Brascher (2014).

a) *obra-to-thema* – qualquer obra pode ter vários *themas* e qualquer tema pode ser assunto de várias obras;

b) *thema-to-nomen* – qualquer *thema* pode ter vários *nomens*, todavia, quando se trata de vocabulários controlados, um *nomen* só pode denominar um *thema*, com o propósito de evitar ambiguidades;

c) *thema-to-thema* – este tipo de relacionamento envolve relações hierárquicas e associativas entre conceitos. As **relações hierárquicas** apresentam-se entre conceitos ou categorias, estabelecidas por meio de superioridade e subordinação, mais comuns em vocabulários controlados, sistemas de cabeçalho de assunto (SILVA, 2014);

d) *nomen-to-nomen* – consiste basicamente em relações de equivalência, ou seja, pode-se dizer que dois *nomens* são equivalentes uma vez que sejam denominações do mesmo *thema* em um vocabulário controlado (SILVA, 2014). No quadro 5, é apresentado de forma sintética os modelos conceituais e suas respectivas entidades:

Quadro 5 - Síntese dos modelos conceituais e suas entidades

MODELO CONCEITUAL	ENTIDADES
FRBR	Obra, Expressão, Manifestação, Item, pessoa, entidade coletiva, família (antes do FRAD), conceito, objeto, evento e lugar
FRAD	Todas as entidades dos FRBR mais as entidades Nome, Identificador, Ponto de acesso controlado, Regra e Agência
FRSAD	Thema e Nomen

RECURSO, DESCRIÇÃO E ACESSO (RDA)

O RDA é alinhado com os modelos conceituais, o que proporciona ao mesmo, uma estrutura que suporte uma infinidade de tipos de conteúdo, tornando ele além de flexível, extensível no registro de dados. O mesmo foi construído sobre os modelos conceituais e padrões internacionais estabelecidos pela IFLA, contemplando as diretrizes dos Princípios Internacionais de Catalogação, os modelos conceituais FRBR, FRAD e o esquema de visualização ISBD que oferece um modo para apresentação dos dados registrados com o RDA (JOINT STEERING COMMITTEE FOR DEVELOPMENT OF RDA, 2010).

Segundo Mey e Silveira (2009), o RDA é composto por 38 capítulos, sendo que o capítulo “0” é basicamente a introdução da norma. O conteúdo do RDA foi desenvolvido na perspectiva de atender as demandas de organização e recuperação da informação como ela se apresenta hoje (OLIVEIRA, 2014).

Ainda segundo Mey e Silveira (2009) o código é composto por 10 seções, este, subdividido em 37 capítulos, além de apresentar 13 apêndices. A norma também faz uso de modelos externos de vocabulário com a *International Standard Organization* – Organização Internacional de Normalização – (ISO), bem como, recomenda o uso de ISBN e ISSN nos registros de identificação.

De acordo com Machado (2014), dentre os objetivos da RDA, destaca-se seu foco em contemplar as necessidades do serviço catalográfico, uma vez que, seu desenvolvimento tem a pretensão de sagrá-lo como uma norma internacional. Sua origem está relacionada sobretudo ao contexto tecnológico, os constantes avanços tecnológicos são prova disso, o que resultou na mudança e aparecimento de diversos suportes, os conteúdos se materializam sob diferentes aspectos, isso implica inclusive na forma de se pensarem os catálogos.

A RDA busca abranger não apenas o ambiente digital, mas também descrever os materiais já presentes na biblioteca, diga-se, os materiais impressos, sem deixar de trazer proposta para os futuros suportes que ainda farão parte da vida dos usuários e que serão exigidos nas UIs. (MACHADO, 2014, p. 9).

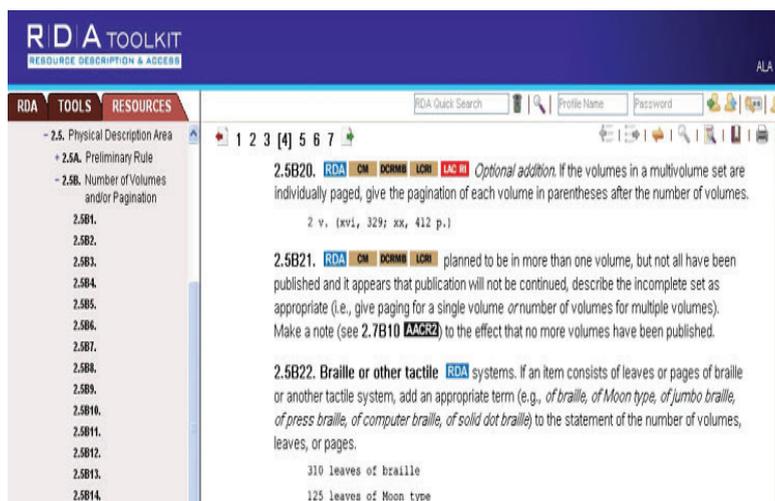
Segundo a *Joint Steering Committee for Development of RDA* (2009), o RDA deve atender a cinco princípios básicos:

- a) generalização – as normas devem contemplar uma infinidade de materiais informacionais, na figura de seus pontos de acesso, nomes e títulos;
- b) especificidade – cada material informacional deve ser tratado individualmente, bem como suas especificidades de conteúdo, nome, título, e a qualquer relação feita entre eles;
- c) ausência de redundância – impedir que hajam repetições;
- d) terminologia – determina que as diretrizes da RDA estejam ligadas aos termos e conceitos dos FRBR e FRAD;
- e) estrutura de referência – evidencia que cada norma e instrução deve se relacionar com normas e instruções gerais, identificando as que são relevantes para seu uso.

Silva (2012), reafirma a projeção da RDA para ambientes digitais, uma vez que o mesmo é capaz de convergir recursos analógicos e digitais, isso por meio de recursos atuais – a exemplo do MARC 21, Dublin Core, e outros que por ventura apareçam – graças à sua flexibilidade e capacidade de extensão no processo de descrição técnica dos mais diferentes tipos de recursos e conteúdo. Esse viés digital da norma é evidenciado

por Oliver (2011), pois, o RDA é projetada para ser uma ferramenta de uso em rede, o RDA Toolkit, conforme figura 13:

Figura 13 – RDA Toolkit



Fonte: RDA Toolkit, 2016.

Voltado para o âmbito digital, o RDA apresenta uma estrutura flexível, e extensível – ora já dito – na perspectiva de contemplar uma infinidade de suportes. A norma, traz consigo algumas diretrizes presentes no próprio AACR2, visando sobretudo, atender os catálogos *online*, resultando em registros mais simplificados e acessíveis (MACHADO, 2014; FUSCO, 2011).

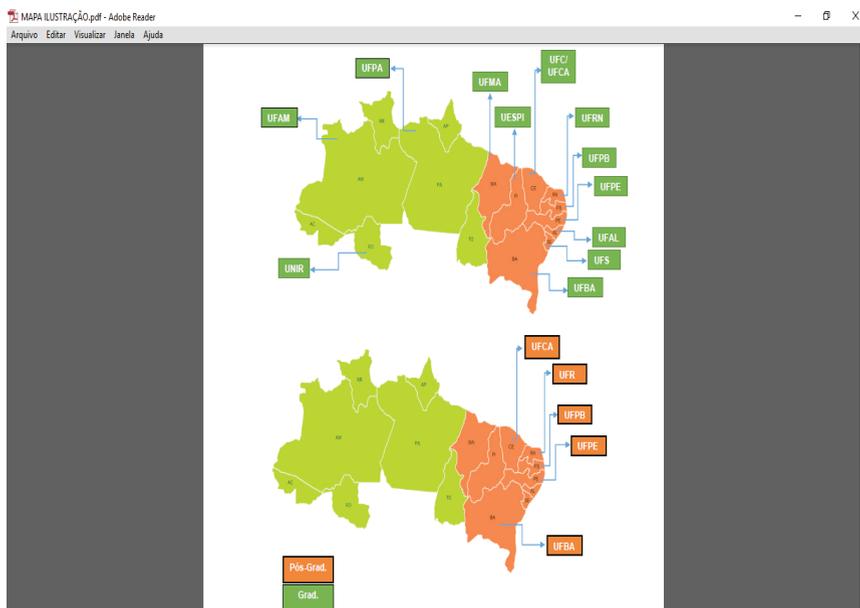
A CATALOGAÇÃO DESCRITIVA REFLETIDA NAS PRODUÇÕES DAS REGIÕES NORTE E NORDESTE

A pesquisa teve como principal objetivo estudar as tendências e limitações nos estudos em catalogação nas regiões Norte e Nordeste, tomando por base as produções científicas dos cursos de Graduação e Pós-Graduação em Biblioteconomia e Ciência da Informação. Sobretudo no intuito de verificar não apenas a presença da catalogação nesses estudos, mas discutir a importância da Representação Descritiva no processo de organização, busca e recuperação da informação, e ver, através dessa cartografia, se as regiões têm acompanhado as movimentações, tanto nacionais, quanto internacionais, quando se falam em tecnologias de representação e descrição da informação.

Os resultados do mapeamento serão apresentados da seguinte forma: primeiro serão apresentados os resultados dos cursos de Pós-Graduação em Biblioteconomia e Ciência da Informação, em seguida dos eventos pesquisados e por fim, os resultados dos periódicos científicos. Ao final se fará um balanço e análise geral com base nas três fontes pesquisadas.

As regiões Norte e Nordeste, segundo dados do E-MEC¹, contam atualmente com 13 cursos de Graduação em Biblioteconomia, nas seguintes IES: UFBA, UFAL, UNIR, UFRN, UFCA, UFPE, UFAM, UFMA, UESPI, UFC, UFPB, UFPA, UFS, no regime de Bacharelado presencial. E segundo dados da CAPES² a região Nordeste conta com 5 cursos de Pós-Graduação em Ciência da Informação (UFBA, UFPB/JP, UFPE, UFC e UFRN) e 2 cursos de Pós-Graduação em Biblioteconomia (UFCA e UFPB/JP). A Região Norte não possui nenhum Curso de Pós-Graduação, tanto de Biblioteconomia, quanto Ciência da Informação. Na figura 14, é apresentada a distribuição geográfica dos cursos:

Figura 14 – Distribuição geográfica dos Cursos de Graduação e Pós-Graduação nas regiões Norte e Nordeste



Fonte: E-MEC, 2016; CAPES, 2016.

¹ <[Http: emec.mec.gov.br/](http://emec.mec.gov.br/)>.

² <<http://www.capes.gov.br/cursos-recomendados>>.

Inicialmente, fez-se necessário identificar se existiam linhas de pesquisa que contemplassem a catalogação descritiva, bem como seus elementos. Nos cursos de Pós-Graduação foram identificadas as seguintes linhas de pesquisa que possuem relação com a catalogação, apresentadas de acordo com os programas e suas respectivas instituições, no quadro 6:

Quadro 6 – Identificação das Linhas de Pesquisa

REGIÃO	PPG	LINHA DE PESQUISA	ÁREAS DE CONCENTRAÇÃO
NE	PPGCI UFPE	Linha de Pesquisa 1 - Memória da Informação Científica e Tecnológica	Dimensões teórico-conceituais da organização e representação da informação e do conhecimento (modelagem conceitual, análise facetada, mapa conceitual, aspectos éticos, instrumentos para a representação (tesauros, taxonomias, ontologias e atuação profissional)
NE	PPGCI UFBA	Linha de pesquisa 1: Políticas e Tecnologias da Informação	Estudos teóricos e aplicados sobre a infraestrutura e políticas de acesso e controle da informação, do documento e das tecnologias intelectuais. Contempla a identificação e o monitoramento de necessidades, assim como a avaliação de padrões de funcionamento e gestão de redes e sistemas de informação. Abrange pesquisas sobre identidade e memória cultural, incluindo o exame de metodologias e estratégias de preservação documental. Envolve ainda o estudo das tendências e dos indicadores de produção e comunicação científica.
NE	PPGCI UFPB	Memória, Organização, Acesso e Uso da Informação	-
NE	PPGCI UFC	LINHA 1 - Representação da Informação e do Conhecimento e Tecnologia	-
NE	PPGB UFCA	Organização e Representação de Informação Documental Produção, Representação, Usos e Impactos da Informação Conceitos, Teorias e Práticas nos Estudos de Memória	-
NE	PPGGI UFRN	Gestão da Informação e do Conhecimento	Gestão de ambientes, de sistemas, de unidades, de serviços, de produtos de informação e de recursos informacionais. Estudos de fluxos, de processos, de uso e de usuários da informação como instrumentos de gestão. Gestão do conhecimento e aprendizagem organizacional no contexto da Ciência da Informação. Marketing da informação, monitoramento ambiental e inteligência competitiva. Redes organizacionais: estudo, análise e avaliação para a gestão. Aplicação das tecnologias de informação e comunicação à gestão da informação e do conhecimento.
N	-	-	-

Fonte: elaborado pelo autor (2016)

Conforme evidenciado no quadro 6, as linhas de pesquisa dos Programas de Pós-Graduação – estes que podem ser *Lato Sensu, Stricto Sensu* ou Técnico - tanto de Ciência da Informação e em Biblioteconomia, contemplam a Catalogação Descritiva, uma vez que, apresentam temáticas voltadas a representação descritiva e representação da informação, modelagem conceitual, taxonomias, ontologias, dentre outros temas. O único programa que não apresentou tanta relação com a temática fora apenas o Programa de Pós-Graduação em Gestão da Informação da UFRN.

A coleta das produções ocorreu também nas Bibliotecas de Teses e Dissertações das IES cujos programas estão vinculados. Utilizou-se a própria ferramenta de busca disponíveis nas BDTDs. Ao todo foram encontradas 40 dissertações.

As produções do Curso de Pós-Graduação da UFCA, não pode ser coletada, pois, por se tratar de uma instituição nova³, ainda não dispõe de uma biblioteca digital. O quantitativo de produções e suas respectivas instituições podem ser melhor visualizadas no quadro 7:

Quadro 7 – Quantitativo de produções e suas respectivas IES

IES	RESULTADOS NA BUSCA	QUANTIDADE QUE TRATAM DE CATALOGAÇÃO
UFBA	0	0
UFPB	26	14
UFPE	12	12
UFC	-	-
UFRN	2	2
UFCA*	-	-
TOTAL	40	28

* Não possui BDTD

Fonte: Mapeamento das BDTDs, 2016

Do total de produções obtidas durante a busca – 40 produções – após a realização da leitura técnica, apenas 28 se tratavam de produções em Biblioteconomia e Ciência da Informação. A UFPB apresentou um número maior de produções, dos 26 resultados na busca, 14 correspondem aos termos estabelecidos para pesquisa. A UFPE apresentou 12 produções acerca das temáticas estabelecidas e a UFRN, apenas duas produções. Quanto a natureza metodológica das produções coletadas, a maioria delas, 18, caracterizam-se como pesquisa empírica, que na perspec-

³ A Universidade Federal do Cariri é uma instituição de ensino superior pública criada em 2013, e sediada na cidade de Juazeiro do Norte, Ceará.

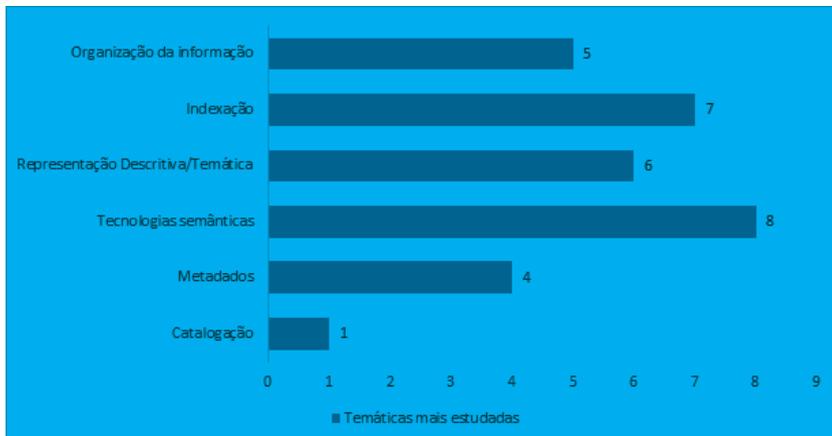
tiva de Prodanov e Freitas (2013), trata-se de um estudo de alto nível de qualificação, de conteúdo original e de profunda reflexão no tratamento de questões teóricas e como complementa Demo (2000, p. 21), é o tipo de pesquisa que se dedica ao estudo da face empírica e fatural da realidade, uma vez que produz e analisa dados por tais meios, oferecendo uma maior concretude de resultados. Das produções restantes, 6 são de natureza qualitativa e 4 Quanti/Qualitativa.

Quanto a metodologia adotada nos estudos coletados, a maioria, 21 produções, configuram-se como pesquisa aplicada, uma vez que, “[...] objetiva gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos. Envolve verdades e interesses locais.” (PRODANOV; FREITAS, 2013, p. 51). 6 produções tratavam de pesquisas teóricas e apenas 1 estudo de caso.

Oliveira (2000, p. 10), ressalta que, a metodologia empregada nas produções da Pós-Graduação passou por mudanças. Segundo a autora, houve uma redução no uso do método quantitativo, e aumento no emprego de metodologias qualitativas. Embora na década de 1990, as metodologias mistas tenham entrado em evidência, há a predominância de procedimentos qualitativos.

Sobre as temáticas mais estudadas nos cursos de Pós-Graduação, estas, podem ser melhor visualizadas no gráfico 1:

Gráfico 1 – Temáticas mais estudadas na Pós-graduação



Fonte: Mapeamento BDTD, 2016

Dentre as temáticas mais estudadas nos resultados das produções dos programas de pós-graduação, estão as tecnologias semânticas e indexação, com 8 e 7 produções respectivamente. Acompanhadas de aborda-

gens acerca das representações descritiva e temática com 6 produções; 5 produções abordaram organização da informação; 4 produções falavam sobre metadados e apenas 1 acerca de catalogação.

Fica evidente o número grande de produções que abordam as Tecnologias semânticas, e pequeno o número de produções acerca da catalogação, partindo disso é possível relacionar com o fato da maioria dos estudos se tratarem de pesquisas teóricas. Os pesquisadores sabem da existência das TS, todavia seu uso ainda não é concretizado, e isso é pontuado na fala de Silva (2011, p. 5), segundo ele, “[...] a aplicação de tecnologias semânticas e métodos de representação de conhecimento, as pesquisas de BI pouco exploram o uso [...]”.

Outrossim, esse número também revela a tendência para o campo da Biblioteconomia e Ciência da Informação, em pesquisas dessa natureza na Pós-Graduação, algo que Silva (2011, p. 143), também assevera, pois,

Verifica-se que algumas iniciativas de BI, inspiradas na Web Semântica, introduzem o uso de tecnologias semânticas e métodos baseados em conhecimento para a exploração dos repositórios da organização. O uso de linguagem natural, similar ao modo familiarizado pelos milhares de usuários dos sistemas de busca na Web, já se revela como uma tendência para a área de BI.

Foram analisadas as produções das regiões Norte e Nordeste nos três principais eventos da área, são eles: o Congresso Brasileiro de Biblioteconomia e Documentação (CBBBD), o Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) e o Seminário Nacional de Bibliotecas Universitárias (SNBU) dos últimos dez anos. No quadro 8 é apresentada a distribuição dos resultados obtidos por evento, a quantidade de produções das regiões Norte e Nordeste.

Quadro 8 – Quantitativo das produções por evento e total das regiões Norte e Nordeste

EVENTO	RESULTADOS NA BUSCA	QUANTIDADE DE PRODUÇÕES NORTE E NORDESTE
CBBBD*	25	3
ENANCIB	185	29
SNBU	374**	16
TOTAL	584	48

* Não foi possível ter acesso aos anais das edições: XXI, XXII, XXIII e XXIV do evento.

** Quantidade total de produções do evento, não apenas referente aos termos pesquisados.

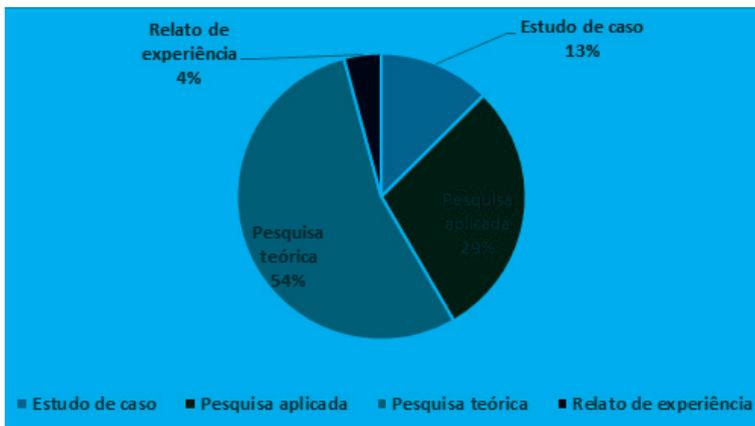
Fonte: Mapeamento dos eventos, 2016

Conforme evidenciado no quadro 13, o evento onde foi possível encontrar o maior número de produções que tratam da catalogação fora o ENANCIB, que de 185 produções obtidas, 29 são provenientes das regiões Norte e Nordeste. O SNBU apresentou 16 produções e o CBBB contou com apenas 3 produções. Cabe destacar que não foi possível ter acesso aos anais de quatro anos seguidos do evento, pois, não foram encontradas páginas, bases ou repositórios que os disponibilizassem, sendo assim, foram analisados apenas os dois últimos anos.

Quanto a natureza metodológica das produções apresentadas nos eventos, a maioria, 39 produções também se tratam de pesquisas empíricas, 6 produções quanti/qualitativa e apenas 3 qualitativas. Até o momento se mantem uma similaridade com as produções dos programas de pós-graduação.

Quanto à metodologia adotada nos estudos, a destruição dos resultados pode ser melhor visualizada no gráfico 2:

Gráfico 2 – Metodologias adotadas nas produções dos eventos



Fonte: Mapeamento dos eventos, 2016

Mediante o gráfico 2, é possível perceber que a maioria das produções – 54% - tratam-se de pesquisas teóricas/revisões de literatura. Em seguida veem as pesquisas aplicadas com 29% dos resultados, e 13% são descritas como estudos de caso, e, por fim, apenas 4% são relatos de experiência.

Quanto as temáticas mais apresentadas nos eventos, a distribuição dos resultados é apresentada no gráfico 3:

Gráfico 3 – Temáticas mais estudadas nas produções de eventos



Fonte: Mapeamento dos eventos, 2016

Dentre as temáticas mais apresentadas estão os trabalhos que abordam organização da informação, com 17 das 48 publicações, em sua maioria, o que vai de encontro à fala de Souza (2013, p. 48), destacando que, “Atualmente, o desafio da Ciência da Informação consiste em contribuir com a sociedade, no que se refere ao acesso à informação por intermédio de sua democratização, socialização, recuperação, organização e disseminação.”

Em seguida, vem as publicações que abordam as representações descritivas e temáticas que somam 11 publicações; produções que falam sobre indexação foram 9; produções que abordam as tecnologias semânticas e catalogação, somaram 6 e 5 trabalhos respectivamente; e por fim, metadados, RDA e MARC21 com 1 produção cada.

Diferente das produções oriundas da pós-graduação, nos eventos já vemos a presença mínima de trabalhos que abordam metadados, RDA e MARC21, isso reflete o despertar dos bibliotecários para esse novo cenário da Catalogação. Os catalogadores brasileiros, de fato tomam propriedade dessas novas ferramentas e de seus impactos nos serviços e sobretudo nas unidades de informação.

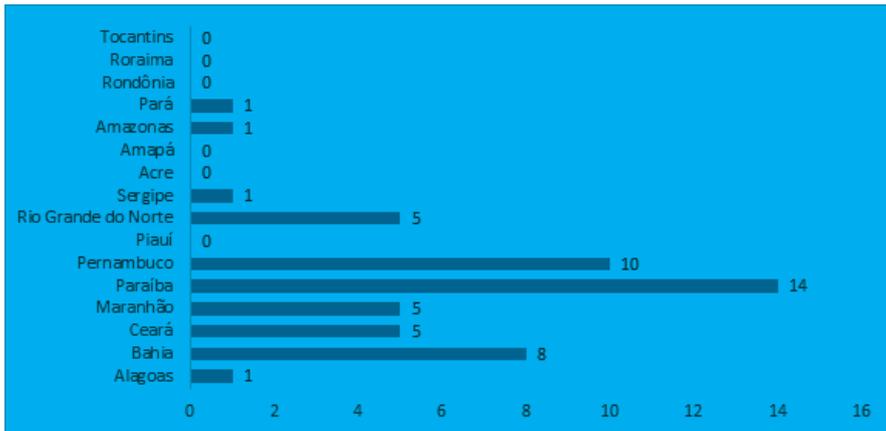
Os últimos acontecimentos no tocante à catalogação, as mudanças advindas das tecnologias e nos próprios ambientes informacionais, despertam a comunidade bibliotecária, a repensar as estratégias de descrição, representação, controle bibliográfico e intercâmbio de informações. Os estudos sobre RDA, vem do estopim dado com os modelos conceituais, o MARC21, e isso é refletido nas produções, a tendência em se estudar os novos instrumentos de descrição.

Nesse sentido, Ferreira e Silva (2013, p. 14), destacam que,

A incorporação de modelos conceituais (FRBR, por exemplo), padrões de metadados (Dublin core, por exemplo) e novo código de catalogação (RDA) estão movimentando esta disciplina e despertando a atenção de estudantes que por muito tempo não entendiam a importância da mesma para a excelência de sua atuação profissional.

Ferreira e Silva (2013, p. 8), justificam os poucos estudos sobre RDA associado ao fato que, “[...] o RDA ainda não consta na ementa das disciplinas analisadas, tal fato não significa que esse assunto não seja tema das aulas, devido às discussões recentes, acredita-se que logo o RDA se formalizará na ementa das disciplinas.” Os estudantes e pesquisadores, começando a ter acesso à novos conceitos, instrumentos e técnicas que estão sendo desenvolvidas e implementadas, a exemplo o próprio RDA. Quanto se partiu para a análise das produções publicadas em eventos, buscou-se identificar também os estados que mais produzem sobre catalogação e tem seus trabalhos publicados nos eventos. Esse ranking pode ser visualizado no gráfico 4:

Gráfico 4 – Estados com o maior número de produções nos eventos



Fonte: Mapeamento dos eventos, 2016

Conforme evidenciado no gráfico 4, os Estados “campeões” em publicações em eventos, cujas produções dizem respeito à catalogação e temas correlatos, é a Paraíba com 14 das 48 publicações encontradas, em seguida, o Estado do Pernambuco que soma 10 produções, e a Bahia com 8 trabalhos apresentados. O Estado do Maranhão aparece nos resultados com 5 produções, empatando com o Ceará e Rio Grande do Norte. Alagoas, Amazonas e Pará contam com apenas 1 publicação, os demais Estados não possuem publicações nos últimos dez anos.

Das duas regiões, a Região Norte, não tem tanta representatividade em estudos sobre Catalogação nos eventos científicos. Essa baixa re-

presentatividade da Região foi estudada por Braga, Hubner e Andretta (2015), os autores destacam que a maioria das produções se concentra em periódicos científicos, evidenciando o baixo número de artigos em eventos, e também, o fato de não haver nenhum programa de Pós-Graduação, na mesma.

E por fim, apresenta-se os resultados do mapeamento realizado nos periódicos científicos com produções do campo da Biblioteconomia e Ciência da Informação. A seleção dos periódicos levou em consideração o indicador *Web Qualis*, conforme explicitado no percurso metodológico. A busca dos artigos foi realizada em suas respectivas páginas online, e utilizou-se o campo de buscas das mesmas.

Ao todo foram mapeados os artigos publicados também no recorte dos últimos dez anos. Ao todo foram mapeados 11 periódicos científicos, e a distribuição dos resultados obtidos pode ser visto no quadro 14:

Quadro 9 – Resultados das produções dos periódicos científicos

PERIÓDICOS/INSTITUIÇÃO/QUALIS			Total	Total N/Ne
Ciência da Informação	IBICT	B1	91	5
DataGramaZero*	IASI	B1	-	-
Em questão	UFRGS	B1	38	1
Encontros Bibli	UFSC	B1	48	3
Informação & Informação	UEL	B1	37	3
Revista Digital de Biblioteconomia e Ciência da Informação	UNICAMP	B1	25	4
Transinformação	PUC-Campinas	A1	50	0
Informação & Sociedade: estudos	UFPB	A1	48	2
Perspectivas em Ciência da Informação	UFMG	A1	42	3
Ponto de Acesso: Revista do Instituto de Ciência da Informação da UFBA	UFBA	B1	50	5
Revista Brasileira de Biblioteconomia e Documentação	FEBAB	B1	22	1
TOTAL			451	27

*Site com problemas, impossibilitando a coleta dos artigos.

Fonte: Mapeamento dos periódicos, 2016

Ao todo foram localizadas 451 produções a partir dos termos buscados nos periódicos listados no quadro 11, desse total, apenas 27 – cerca de 6% - são produções nativas das regiões Norte e Nordeste. Um número muito pequeno se comparado ao total de produções identificadas. O periódico que contém o maior número de artigos oriundo das duas regiões é a revista *Ciência da Informação* e a *Ponto de Acesso: Revista do Instituto de Ciência da Informação da UFBA*, ambas com 5 artigos cada. A *Transinformação* não apresentou nenhum resultado, assim como, não foi possível coletar as publicações da *DataGramZero*, tendo em vista que o site se encontra com problema, impossibilitando a mesma.

Quanto a natureza metodológica, a maioria dos estudos, 24 produções, se caracterizam como pesquisa empírica, 2 estudos tratam-se de estudos qualitativos e 1 quanti/qualitativo. Quanto as metodologias adotadas nos estudos, a maior parte utilizou da pesquisa teórica, 16 das 27 produções; 9 são estudos que tratam de pesquisas aplicadas e apenas 1 estudo de caso.

No gráfico 5, são apresentadas as temáticas predominantes nas produções identificadas nos periódicos:

Gráfico 5 – Temáticas mais estudadas nos periódicos



Fonte: Mapeamento dos eventos, 2016

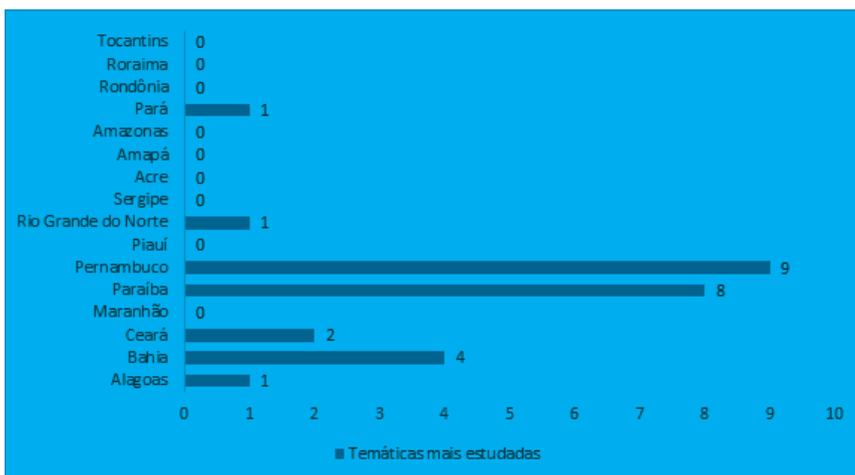
Conforme evidenciado no gráfico 5, a maioria dos trabalhos abordam as representações descritiva e temática, totalizando 10 produções; 8 produções tratam da organização da informação; indexação e tecnologias semânticas também são estudadas, com 6 e 5 produções respectivamente, e, apenas uma produção abordou metadados. Tais informações vão de encontro a fala de Oliveira (2005, p. 11, grifo nosso), ao falar das tendências temáticas em produções científicas da área, segundo a autora,

Observa-se que os estudos sobre Usuários, transferência e uso da informação e da biblioteca, sobre Processamento e recuperação da informação (entrada, tratamento, armazenamento, recuperação e dis-

seminação da informação) e sobre Gerência de serviços e unidades de informação foram os assuntos mais pesquisados e, portanto, com maior volume de produção.

Assim como evidenciado no mapeamento das produções científicas apresentadas em eventos, também são ranqueados os estados com mais publicações nos periódicos, conforme gráfico 6:

Gráfico 6 – Estados com o maior número de produções em periódicos científicos



Fonte: Mapeamento BDTD, 2016

Conforme evidenciado no gráfico 6, Pernambuco larga na frente com a maior quantidade de produções em periódicos, totalizando 9 publicações, seguido da Paraíba com 8 publicações. Bahia e Ceará contam com 4 e 2 produções respectivamente, e Alagoas, Pará e Rio Grande do Norte com 1 publicação cada.

Fazendo um panorama geral, tomando por base todos os resultados encontrados, e após a identificação das produções oriundas das regiões Norte e Nordeste, de um montante de 1075 publicações, as duas regiões acumulam apenas 103 produções – entre dissertações e artigos, tanto de periódicos científicos da área, quanto de eventos – cerca de 10% do total de produções, é de certa forma um número pequeno, uma vez que, as duas regiões possuem 13 cursos de graduação e 7 cursos de pós-graduação.

A partir do mapeamento é possível apontar os Estados da Paraíba e Pernambuco como “potências” nos estudos em catalogação quando se fala em Norte e Nordeste, de longe os dois estados, possuem o maior número de publicações nos três eixos, onde a Paraíba soma 36 produções

e Pernambuco 31 produções. Fazendo uma breve relação com os demais Estados da Federação, a quantidade de produções acerca da catalogação descritiva nas regiões Sul e Sudeste são maiores que nas regiões Norte e Nordeste, sobretudo na Região Norte que possui um número bem menor de produções.

A baixa representatividade da Região Norte no tocante à produções, não apenas se tratando de estudos em Catalogação, mas em Biblioteconomia e Ciência da Informação no todo, fora abordado no estudo de Braga, Hubner e Andretta (2015). Segundo os autores, há pesquisas sendo desenvolvidas na região, no campo da Ciências Sociais Aplicadas, e os resultados, em sua maioria são publicados em periódicos científicos, sobretudo, é ressaltado que, há falta de oferta de Pós-Graduação em Ciência da Informação.

Nesse sentido, Braga, Hubner e Andretta (2015, p. 3), dizem que,

Considera-se conveniente, para o desenvolvimento da Ciência da Informação na região Norte a oferta de Programas de Doutorado Interinstitucional (DINTER) para os docentes, assim como a constituição de parcerias intra e inter departamentais e consolidação de uma agenda de pesquisa.

E quando comparada a Região Norte à Região Nordeste, a segunda se mostra mais forte nos estudos de catalogação, seja pela quantidade produções nos três eixos (programas de Pós-Graduação, eventos e periódicos científicos), mas sobretudo, por ser sede de 10 dos 13 treze cursos de Graduação e dos 7 cursos de Pós-Graduação em Biblioteconomia e Ciência da Informação, refletindo no número de produções.

Segundo Modesto (2007), a Catalogação sempre fora explorada em menor grau pela Pós-Graduação, isso ainda na década de 1990, e isso é evidenciado na identificação das linhas de pesquisa dos programas da Região Nordeste (cabe destacar que a Região Norte não possui nenhuma pós-graduação em Biblioteconomia e Ciência da Informação ativos). A catalogação não possui nenhuma linha de pesquisa exclusiva em nenhum dos programas mapeados, todavia, são contempladas nos temas correlatos à cada linha, e no caso da PPGB da UFCA, que o termo "Representação" está em destaque. Isso reflete significativamente na quantidade de produções.

Nas produções identificadas nos principais eventos de Biblioteconomia e Ciência da Informação, a catalogação ainda se apresenta timidamente. A maioria das produções são revisões de literatura. O caráter prático do processo de descrição e representação da informação não é posto em destaque nos eventos. O mesmo cenário é observado nas publica-

ções periódicas, um número pequeno de publicações e em sua maioria revisões de literatura.

No estudo recente de Freire e Souza (2010, p. 122), ao realizarem um mapeamento temática da produção da Revista Brasileira em Ciência da Informação e Biblioteconomia, também tomando por base a análise de assunto, evidenciaram os temas recorrentes na “Categoria Representação”, e segundo os autores, os temas mais comuns, foram: metadados, linguagens documentárias, mapas conceituais, Tesouros, Sistema de Recuperação da Informação. Percebe-se então, que os estudos não abordam as temáticas mais recentes no tocante à catalogação.

Quanto as temáticas abordadas nos três eixos (pós-graduação, eventos e periódicos científicos), destacam-se estudos que abordam organização da informação, indexação e representação da informação. Temáticas mais atuais como Tecnologias Semânticas, RDA, FRBR são pouco explorados nas produções das regiões. Evidenciando assim, um atraso quando se fala em catalogação, uma vez que, ferramentas mais atualizadas não são utilizadas ou não são abordadas nos trabalhos.

Ao analisar o sumário do último Encontro Internacional de Catalogação e Encontro Nacional de Catalogação⁴, observa-se um certo equilíbrio nos estudos de catalogação. No sumário, pode-se observar estudos que envolvendo não apenas os aspectos teóricos da Catalogação – a exemplo o trabalho intitulado “O formato MARC e o RDA: tempo de mudanças”, “*Resource Description and Access* (RDA) e as mudanças na catalogação” – mas também, aspectos metodológicos – quando se fala em inquietações no ensino de representação descritiva – e, sobretudo práticos – quando se apresentam relatos de experiência, “Controle de autoridades de nomes pessoais: relato de experiência do SIBI/UFSC”, além de estudos que evidencia a crescente das tecnologias, e uso de metadados.

Quando tomamos por base o maior e mais representativo evento do campo da catalogação – o Encontro Nacional de Catalogadores e o Encontro Internacional de Catalogação – é possível perceber o não alinhamento das regiões com as discussões nacionais, pois, fazendo um breve levantamento das temáticas discutidas nos últimos anos, entre elas, os Requisitos Funcionais para Registros Bibliográficos (FRBR), a representação descritiva em tempos digitais, as Tecnologias Semânticas que estão em evidência, até mesmo o próprio RDA – este que por sua vez apresentou apenas uma produção no total dos três eixos – destaca que as regiões não tem de fato acompanhado os passos que a Catalogação tem dado, tanto no cenário nacional, quanto internacional.

4 Disponível em: <<http://www.telescopium.ufscar.br/index.php/eic-enacat/eic-enacat/schedConf/presentations>>.

O último encontro do Grupo de Estudos e Pesquisas em Catalogação (GepCat)⁵ denotam esse novo cenário da catalogação nacional. Temas como Tecnologia Semântica RDF como instrumento de descrição, FRAD, RDA, ISBD, nos fazem repensar a catalogação feita nas regiões, e, recoloca os bibliotecários em um novo momento.

Mesmo lentamente, a área cresceu tanto quantitativa quanto qualitativamente, salientando-se essa melhoria, sobretudo, no que diz respeito a dois aspectos considerados fundamentais pela avaliação: a inserção das pesquisas no campo da Ciência da Informação e a produção bibliográfica resultante deste esforço. (OLIVEIRA, 2005, p. 12).

Ferreira e Silva (2013), destacam que de fato a Catalogação dispõe de poucos grupos de pesquisa, isso se justifica pelo fato de ser uma área muito específica, uma vez que, a Biblioteconomia e Ciência da Informação dispõem de diversos eixos temáticos, que possuem praticamente o mesmo número de grupos de pesquisa. Todavia, o interesse pela representação descritiva não é isolado, e, é importante que pesquisas nesta área se desenvolvam. De fato, a Catalogação Descritiva, suscita inúmeras discussões, sobretudo, mediante o impacto causado pelas novas tecnologias nas práticas da representação descritiva.

É fundamental que a Catalogação garanta seu espaço não apenas no Norte e Nordeste, mas que se consolide por meio do desenvolvimento teórico, e, sobretudo pela prática profissional, de modo a empregar as diversas ferramentas no processo de representação e descrição da informação. A validade deste mapeamento se dá pela importância de estudos desta natureza, uma vez que não apenas se analisam as produções, mas que estas análises sirvam como parâmetro e indicar científico para a área.

REFERÊNCIAS

ALVES, Rachel Cristina Vesú. **Metadados como elementos do processo de catalogação**. 2010. 134f. Tese (Doutorado em Ciência da Informação) Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

ALVES, Rachel Cristina Vesú; SANTOS, P. V. A. da C. **Metadados no domínio bibliográfico**. Rio de Janeiro: Intertexto, 2013.

BDTD. (Mapeamento BDTD, 2016)? Qual página foi encontrada as informações do Gráfico?

BEZERRA, Darlene Alves. **O modelo FRBR e a busca de semântica na catalogação e recuperação de informação em ambientes**. 185f. 2013.

⁵ Maiores informações podem ser obtidas no seguinte endereço: <<http://gepcat.blogspot.com.br/>>.

Dissertação (Mestrado em Ciência da Informação) - Universidade Federal Fluminense, Niterói, 2013.

BEZERRA, Darlene Alves; DE SOUZA, Elisabete Gonçalves. Os FRBR e a descrição de metadados em meio digital. **EDICIC**, v. 1, n. 4, p. 327-341, 2011.

BRAGA, Aurineide Alves; HUBNER, Marcos Leandro Freitas; ANDRETTA, Pedro Ivo Silveira. **A produção bibliográfica nos departamentos de Ciência da Informação da região norte do Brasil**: aproximações e perspectivas. 2015. Disponível em: http://siscone.com.br/Uploads/CBBDD15/Trab14400187320150329_000000.pdf. Acesso em: 19 jun. 2021

CAPES. **Cursos recomendados/reconhecidos**: região 2016 Disponível em: https://sucupira.capes.gov.br/sucupira/public_consultas/coleta/programa/quantitativos/quantitativoRegiao.jsf;jsessionid=lpC19tcusCVdbQWNHksjYjWE.Sucupira-213. Acesso em: 08 abr, 2021.

CASTRO, Fabiano Ferreira de; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Uso das tecnologias na representação descritiva: o padrão de descrição bibliográfica semântica MacOnt Initiative nos ambientes informacionais digitais. **Ci. Inf.**, v. 38, n. 1, jan./abr. 2009.

CORRÉA, Rosa Maria Rodrigues. **Catálogo descritivo no século XXI**: um estudo sobre o RDA. 2008. 75f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2008.

E-MEC, **Cadastro Nacional de Cursos e Instituições de Educação Superior**. 2016. Disponível em: <http://emec.mec.gov.br>. Acesso em: 13 ago. 2021

FERREIRA, Valéria Alves; SILVA, Marcia Regina. Representação descritiva no Brasil: ensino e pesquisa. In: Encontro Internacional de Catalogadores, 9., 2013, Rio de Janeiro. **Anais...** Rio de Janeiro: Fundação Biblioteca Nacional do Brasil, 2013.

FREIRE, Isa Maria; SOUZA, Alexandre Pereira. Revista Pesquisa Brasileira em Ciência da Informação e Biblioteconomia-PBCIB: um mapeamento temático da produção científica à luz da análise de conteúdo/Revista Pesquisa Brasileira em Ciência da Informação e Biblioteconomia-PBCIB: um mapeamento. **Informação & Informação**, v. 15, n. 2, p. 110-128, 2010.

FUSCO, Elvis. **Aplicação dos FRBR na modelagem de catálogos bibliográficos digitais**. São Paulo: Cultura Acadêmica, 2011.

FUSCO, Elvis. **Modelos conceituais de dados como parte do processo de catalogação**: perspectiva de uso dos FRBR no desenvolvimento de

catálogos bibliográficos digitais. 2010. 249f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

FUSCO, Elvis; SANTOS, Plácida L. V. A. C. A modelagem de dados no processo da catalogação sob a perspectiva do uso dos FRBR. In: Seminário em Ciência da Informação, 3., 2009, Fortaleza. **Anais...**Fortaleza: UFC, 2009.

GOMES, Maria Yêda Falcão Soares de Filgueiras. Tendências atuais da produção científica em Biblioteconomia e Ciência da Informação no Brasil. **DataGramaZero - Revista de Ciência da Informação**, v. 7, n. 3, jun. 2006.

GRÁCIO, José Carlos Abbud. **Metadados para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade**. 2002. 127f. 2002. Tese de Doutorado. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.

IFLA. **Declaração de Princípios Internacionais de Catalogação**. 2009. Disponível em: www.ifla.org/VII/s13/icp. Acesso em: 08 abr. 2021

IFLA. **Functional requirements for bibliographic records: final report**. München: K. G. Saur, 1998.

IFLA. **Functional Requirements for Subject Authority Data (FRSAD): a Conceptual model**. Berlin: IFLA, 2010.

JOIN STEERING COMMITTEE FOR DEVELOPMENT OF RDA. **RDA - Resource Description and Access: objectives and principles**. 2009. Disponível em: <http://www.rda-jsc.org/archivedsite/docs/SRDAobjectivesrev3.pdf>. Acesso em: 27 maio. 2021

JOINT STEERING COMMITTEE FOR DEVELOPMENT OF RDA. **Frequently asked questions**. [S.l.]: [S.n.], 2010.

LOURENÇO, C. A. L. **Modelagem de dados como ferramenta de análise de padrões de metadados em bibliotecas digitais: o padrão de metadados brasileiro para teses e dissertações segundo o modelo entidade-relacionamento**. 2005. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

MACHADO, Raildo de Sousa. Recurso, Descrição e Acesso - RDA: breve descrição. In: Seminário Nacional de Bibliotecas Universitárias, 18., 2014, Belo Horizonte. **Anais...** Belo Horizonte: SNBU, 2014.

MELO, Maria Antônia Fonseca; BRÄSCHER, Marisa. Requisitos Funcionais para Dados de Autoridade Assunto (FRSAD): entidades, atributos e relacionamentos. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 12, n. 2, p. 102-119, 2014.

MEY, Eliane Serrão Alves; SILVEIRA, Naira Christofoletti. **Catálogo no plural**. Brasília: Briquet de Lemos, 2009.

MODESTO, Fernando. **A biblioteca e os recursos informacionais na era digital**: copyright e licenciamentos. 2005. Disponível em: https://www.ofaj.com.br/colunas_conteudo.php?cod=228. Acesso em: 20 jul. 2021

MODESTO, Fernando. Panorama da catalogação no Brasil: da década de 1930 aos primeiros anos do século XXI. In: Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação, 22., 2007, Brasília. **Anais...** Brasília: CBBBD, 2007.

MORENO, Fernanda Passini. **Requisitos Funcionais para Registros Bibliográficos - FRBR**: um estudo no catálogo da Rede Bibliodata. 202f. 2006. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília, Brasília, 2006.

MORENO, Fernanda Passini; MÁRDERO ARELLANO, Miguel Ángel. Requisitos funcionais para registros bibliográficos - FRBR: uma apresentação. **Revista Digital de Biblioteconomia e Ciência da Informação**, v. 3, n. 1, 2005.

OLIVEIRA (2000) tendências temáticas em produções científicas da área (referência não encontrada)

OLIVEIRA, Marcelo Augusto Cardoso de. **RDA**: utilização do novo código de catalogação pelas bibliotecas das universidades federais brasileiras. 69f. 2014. Monografia (Graduação em Biblioteconomia) - Faculdade de Informação e Comunicação, Universidade Federal de Goiás, Goiânia, 2014.

OLIVER, Chris. **Introdução à RDA**: um guia básico. Brasília, DF: Briquet de Lemos, 2011.

PACHECO, Elizabete Moraes. **Entidades e relacionamentos do FRBR**: uma discussão a partir da obra *Ulysses* de James Joyce. 70f. 2014. Monografia (Graduação em Biblioteconomia) - Universidade Federal do Rio Grande, Rio Grande, 2014.

PAES, D. M. B.; TABOSA, H.; PINTO, V. B. Uso de ferramentas tecnológicas na representação temática e descritiva da informação: relato de experiência na iniciação à Docência na Unidade Curricular de Processamento da Informação. In: Encontro Regional de Estudantes de Biblioteconomia,

Documentação, Ciência da Informação e Gestão da Informação, 16., 2011, São Luís. **Anais...** São Luís: UFMA, 2011.

PIRES, D. Uso do Dublin Core na descrição de obras raras na web: a coleção da Biblioteca Brasileira Digital. In: Encontro Nacional de Catalogadores, 1., 2012, Rio de Janeiro. **Anais...** Rio de Janeiro: ENACAT., 2012.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do Trabalho Científico: métodos e técnicas da pesquisa e do trabalho lêmico**. 2. ed. Novo Hamburgo, RS: Universidade Feevale, 2013.

RDA TOOLKIT. **Resource description & access**. 2016. Disponível em: <http://ess.rdatoolkit.org>. Acesso em: 19 jun. 2021.

SILVA, Dhiogo Cardoso da. **Uma arquitetura de business intelligence para processamento analítico baseado em tecnologias semânticas e em linguagem natural**. 161f. 2011. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2011.

SILVA, Diego Maradona Souza da. **Representação descritiva na contemporaneidade: o caso dos Requisitos Funcionais para Registros Bibliográficos (FRBR) e da Recursos: Descrição e Acesso (RDA)**. Monografia (Graduação em Biblioteconomia) - Universidade Federal do Rio Grande do Norte, Natal, 2012.

SILVA, Luciana Candida; BAPTISTA, Dulce Maria. Entre a teoria e a prática no ensino do formato MARC 21: a metodologia da Universidade Federal de Goiás. In: Encontro Internacional de Catalogadores, 9.; | Enacat-Encontro Nacional de Catalogadores, 2., 2013, Rio de Janeiro. **Anais...** Rio de Janeiro: Fundação Biblioteca Nacional do Brasil, 2013.

SILVA, Michele Mello da. **O modelo conceitual “Requisitos Funcionais para Dados de Autoridade de Assunto” (FRSAD): estrutura, aplicação e relação com os modelos FRBR e FRAD**. 65f. 2014. Monografia (Graduação em Biblioteconomia) - Universidade Federal do Rio Grande, Rio Grande, 2014.

SOUSA, Ana Claudia Medeiros de. Temáticas das dissertações da Pós-Graduação em Ciência da Informação da UFPB - 2008/2012. **RACIn**, João Pessoa, v. 1, n. 2, p. 37-53, jul./dez. 2013.

SOUZA, Marcia Izabel Fugisawa; VENDRUSCULO, Laurimar Gonçalves; MELO, Geane Cristina. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ci. Inf**, v. 29, n. 1, jan./abr. 2000.



CATALOGAÇÃO COOPERATIVA NO SISTEMA DE BIBLIOTECAS DA ELETROBRÁS/ELETRONORTE

Jousiane Leite Lima

INTRODUÇÃO

A biblioteca constitui-se como organização que permeia questões envolvendo a preservação, representação e disseminação da informação e do conhecimento. Nesta perspectiva, esta organização desenvolve-se suas atividades de acordo com os avanços tecnológicos, incluindo desde a utilização do computador com a automação de bibliotecas em meados dos anos 60, até a comunicação em redes, destacando-se a Internet.

Essa evolução tecnológica implica em trabalhos ágeis e de acesso em tempo real às informações disponibilizadas na Internet, permitindo, desta forma, o compartilhamento de dados e informações entre bibliotecas.

A catalogação cooperativa é um exemplo de atividade exercida dentro de uma unidade de informação, a qual foi aprimorando-se com os avanços da Tecnologia de Informação e Comunicação, e é exigida a utilização de instrumentos de padronização internacional, como o Anglo-American Cataloging Rules – Second Edition (AACR2) Código de Catalogação Anglo-Americano, o formato Machine Readable Cataloging (MARC) e os protocolos de comunicação como o Z39.50 e a norma de intercâmbio de dados ISO 2709.

A catalogação acompanha as evoluções tecnológicas na tentativa de definir ferramentas mais eficazes para a representação dos recursos informacionais, disponíveis também agora em meio eletrônico. Esses recursos requerem métodos de representação mais específicos que os métodos tradicionais, ou seja, necessitam de novas formas de representação que atendam a suas características.

CATALOGAÇÃO COOPERATIVA: conceitos, histórico e instrumentos de padronização

Ao refletir sobre a história da humanidade, percebe-se que os homens possuem uma característica diferenciada em relação aos outros animais: a racionalidade. Dessa forma, eles criaram várias formas de sobrevivência de acordo com as suas necessidades e ambiente em que viviam, até chegar ao estágio atual da sociedade. Outra característica que foi fundamental para a sobrevivência humana chama-se cooperação.

Assim como o próprio ser humano, os serviços informacionais foram modificando-se com o passar dos anos, as atividades biblioteconômicas tornaram-se precisas e operacionalizáveis. Nessa direção está o trabalho de catalogação.

Desse modo, Vetter e Araújo (2012, p. 2) expõem que a catalogação consiste no processo de descrição do documento, para identificá-lo, entre tantos outros, mantendo suas características individuais e ao mesmo tempo possibilitando sua relação com outros documentos.

Para tanto, durante muitos anos, a catalogação preocupou-se com a preparação e produção de fichas catalográficas pelos bibliotecários, no qual exigia muito trabalho e tempo, podendo considerar um fator preocupante nessa atividade por parte dos mesmos.

A catalogação foi desenvolvendo-se juntamente com o campo de automação de bibliotecas e com isso “[...] as bibliotecas [e os bibliotecários] perceberam que não poderiam adquirir e processar materiais pelos sistemas tradicionais e que automatizar poderia também ajudar a controlar custos” (ORTEGA, 2002, p. 35).

Ortega (2002) continua ressaltando que a catalogação sendo uma das operações mais caras em bibliotecas, e materiais comuns existia em locais distintos, o registro de um material realizado por uma biblioteca poderia ser aproveitado por várias outras. Pode-se afirmar então que essa ideia exprime a questão de compartilhamento e intercâmbio de dados, e de catalogação cooperativa.

As bibliotecas são entidades que de acordo com Mey e Silveira (2009, p. 8) precisam cooperar entre si e trabalhar integradamente em várias atividades. Quanto maior a integração, maiores se tornam os acervos, mais rapidamente se faz a representação bibliográfica e menores resultam os custos.

Entretanto, o trabalho cooperativo entre bibliotecas significa o “uso de uma linguagem comum, de padrões” (MEY; SILVEIRA, 2009, p. 9).

Dessa forma, Lopes (2010, p. 43) enfatiza que

A catalogação cooperativa se dá quando um grupo de bibliotecas se une, procurando eliminar a duplicação de esforços e o desperdício de recursos. Nesta perspectiva, forma-se um catálogo coletivo, onde o item que já foi catalogado por uma biblioteca não precisa mais ser catalogado pelas outras. Antigamente, as bibliotecas trabalhavam isoladamente e, muitas vezes, duas ou mais bibliotecas de uma mesma cidade catalogavam o mesmo livro. Isso aumentava significativamente as despesas com recursos humanos. Um dos benefícios que a formação de redes de informação trouxe é a catalogação cooperativa. [...] O avanço da tecnologia da informação possibilitou o compartilhamento de informações catalográficas de modo instantâneo.

Assim, a catalogação cooperativa permite que em uma rede de bibliotecas evite-se catalogar um documento várias vezes, na medida em que as informações em diferentes formatos e suportes, a cada instante, aumentam devido aos avanços tecnológicos. Nesse sentido, o bibliotecário economiza tempo e trabalho intelectual ao fazer a representação descritiva de um documento.

Barbosa (1978, p. 73) também afirma que

O trabalho isolado há muito perdeu sua razão de ser. A cooperação aplicada à catalogação vem transformando-a, pouco a pouco, numa disciplina revestida de novas características. Anteriormente, a função da catalogação era apenas a de servir como veículo de registro das coleções; sua redação trabalhosa e complicada tornava-a uma tarefa quase que indesejável. Hoje, sistematizada e adaptada às técnicas modernas, alia à sua função anterior, a de servir também como veículo de transmissão da informação.

Atualmente, a catalogação cooperativa favorece o intercâmbio de informações entre bibliotecas, embora somente em meados do século XX, a catalogação cooperativa tornou-se algo tangível.

Campello (2006, p. 68) relata que o pioneiro da catalogação cooperativa foi o bibliotecário Charles Jewett (1816-1868), que em 1850, em um encontro da *American Association for the Advancement of Science* propôs que a biblioteca *Smithsonian Institution* (Estados Unidos) onde ele trabalhava fosse a Biblioteca Nacional. Funcionaria como uma central de catalogação nacional, porém essa iniciativa não teve incentivo.

Em 1901 a Library of Congress (LC), concretizou a ideia de Charles Jewett em relação à catalogação cooperativa: “a impressão e venda de fichas. Ao invés de cada biblioteca elaborar a própria catalogação de seus livros, a LC passou a vender suas fichas impressas” (MEY; SILVEIRA, 2009, p. 73).

Após algum tempo, a LC além de imprimir e vender suas fichas catalográficas, “passou a receber registros de outras bibliotecas e deixou de ser centralizado para se tornar cooperativo” (CAMPELLO, 2006, p. 68).

Incidiram assim algumas iniciativas relevantes no âmbito de legitimar a catalogação cooperativa como prática biblioteconômica em vários países do mundo, inclusive no Brasil, a fim de que buscassem uma padronização ao representar informações.

No Brasil, a catalogação cooperativa teve impulso em 1942, com a bibliotecária Lydia de Queiroz Sambaquy, na biblioteca do Departamento Administrativo do Serviço Público (DASP), que criou o Serviço de Intercâmbio de Catalogação (SIC) (CAMPELLO, 2006, p. 69).

Segundo Campello (2006, p. 69) qualquer biblioteca poderia participar do SIC, enviando suas fichas catalográficas, que seriam revistas, impressas e distribuídas as bibliotecas cooperantes, porém a falta de conhecimento dos bibliotecários sobre catalogação da época implicou em um maior cuidado na revisão das fichas antes de serem impressas.

Outra iniciativa foi, em 1947, com a Fundação Getúlio Vargas (FGV), pois

Decidiu, por iniciativa própria, desenvolver um sistema de automação baseado em um padrão proposto e já suficientemente validado por instituições representativas da Biblioteconomia nacional, primeiramente para uso interno, mas voltado, desde sua concepção, para servir de infraestrutura para a formação de uma rede nacional de catalogação cooperativa que, esperançosamente, viria a se estabelecer em torno deste sistema. Esta expectativa foi confirmada na prática, e os anos subsequentes assistiram a uma adesão gradual e crescente à rede Bibliodata de uma parcela expressiva de instituições de diversos estados brasileiros (VASCONCELOS 1996, p. 2).

A catalogação cooperativa foi, então, iniciada no Brasil já em moldes aceitos internacionalmente, porém, com diversos reparos a partir dos interesses específicos das instituições que primeiro a utilizaram, como por exemplo, a FGV (GRAVINA, 2012).

Para tanto, Mey e Silveira (2009, p. 75) expõem que nos anos 1970 e 1980, o mundo caminhava para a padronização, pressionado pelos avanços tecnológicos, e todos os códigos iriam chegar a pontos comuns.

No entanto, segundo Silveira e Tálamo (2009, p. 109) mesmo que isoladamente, os bibliotecários estipulavam padrões para a representação de documentos, sendo que com as mudanças na sociedade e na forma de produção de conhecimento, a padronização deixou de ser realizada em nível local e se expandiu para o nível internacional.

Vale destacar que o bibliotecário que pratica e utiliza a catalogação cooperativa exerce grande responsabilidade, pois se o trabalho não estiver condizente com os padrões estabelecidos internacionalmente, serão afetadas todas as bibliotecas cooperantes e a recuperação da informação por parte dos usuários. Por isso a importância de haver treinamentos e discussões sobre as políticas de catalogação dessas bibliotecas e quem sabe avaliação desse serviço.

Nesta perspectiva, Santos (2005, p. 5) considera que o trabalho realizado juntamente com outros profissionais tem maior probabilidade de ser realizado com um grau mais elevado de precisão, pois trabalhando juntos, os bibliotecários dão-se a oportunidade de trocar experiências e crescer profissionalmente juntamente com os demais profissionais da área.

Há vantagens das bibliotecas utilizarem a catalogação cooperativa na sua rotina, pois permite a redução no tempo ao descrever um item, bem como o trabalho cooperativo entre os profissionais. Por outro lado, ainda existem profissionais que não assimilaram esse trabalho cooperativo, pois preferem fazer seu serviço individualmente e de forma tradicional.

No entanto, com a introdução da automação de bibliotecas fez com que a catalogação cooperativa avançasse, sendo que

Passados quase 50 anos, percebe-se mudança significativa no contexto da atividade catalográfica, representada principalmente pelo desenvolvimento de serviços informatizados de criação e processamento de dados bibliográficos e pelo aparecimento de grandes bases de dados, nacionais e internacionais, de catalogação cooperativa (CAMPELLO, 2006, p. 61).

Esse “novo” ambiente favoreceu o campo técnico das práticas biblioteconômicas, pois outrora o bibliotecário levava tempo para tratar uma informação, hoje com o uso da tecnologia da informação é mais simples o usuário recuperar uma informação tratada, atualizada em tempo hábil.

A automação adentrou as bibliotecas para otimizar os serviços já existentes, assim como a criação de novos. Campello (2006, p. 69) identifica que a automação possibilitou o aparecimento de redes de bibliotecas unidas por interesses comuns, como as bibliotecas universitárias, uma vez que essas instituições anualmente recebem milhares de materiais, e se o bibliotecário fosse catalogar item por item levaria muito tempo para tratar a informação e conseqüentemente o usuário teria acesso ao item tardiamente.

Com a automação de bibliotecas, “a catalogação cooperativa, juntamente com a padronização dos registros bibliográficos, auxilia este processo uma vez que alguém já pode ter feito a catalogação do documento, resta apenas, compartilhar o registro, economizando tempo e trabalho” (SANTOS, 2005, p. 4).

A automação de bibliotecas determinou grande impulso na catalogação cooperativa diminuindo os custos com o processamento técnico e aumentando a padronização da descrição. A existência de redes de catalogação cooperativa possibilita a criação de bancos de dados bibliográficos que também fornecem registros já existentes às suas cooperantes.

Dessa forma, com os avanços tecnológicos, houve a expansão do serviço de catalogação, necessitando da criação de instrumentos de padronização internacional para este tipo de atividade, permitindo a cooperação entre bibliotecas, como:

a) AACR2 (Anglo-American Cataloging Rules – Second Edition) - padrão internacional de descrição dos acervos bibliográficos a partir do qual foi desenvolvido o Formato MARC 21 (ROMANI, 2006).

O AACR2 tem como objetivo a normalização da catalogação a nível internacional, subsidiando o tratamento da informação. Faz uso do sistema de pontuação e a catalogação pode ser realizada pelo suporte físico da obra.

b) MARC - padrão para catalogação legível por computador, onde os dados catalogados são processados por este e colocados de maneira compreensível pela máquina, no qual são identificados os elementos de forma explícita pra fins de manipulação.

De acordo com Romani e Borszcz (2006, p. 114)

MARC21 (Machine Readable Cataloging): é o registro de catalogação legível por máquina, ou seja, é um padrão internacional de descrição automatizada dos acervos bibliográficos utilizado pelos principais programas de bibliotecas. Como é um formato de intercâmbio, destina-se a padronizar a representação dos itens bibliográficos em sistemas informatizados.

Barbosa (1978) afirma que o MARC começou como projeto piloto que distribuía os registros da Library of Congress, de modo claro pelo computador, para as bibliotecas participantes. O Projeto Piloto teve como objetivo testar os benefícios e problemas da produção centralizada de registros catalográficos legíveis por computador e posteriormente distribuir às bibliotecas usuárias, as fichas produzidas. O projeto piloto foi idealizado na década de 1960, aplicado em 16 bibliotecas que passaram a usar as fitas magnéticas MARC como entrada para produzir catálogos

impressos em forma de livros, fichas para catálogos, etiquetas, listagens bibliográficas parciais etc. Programado para terminar em junho de 1967, estendeu-se até junho de 1968 pela necessidade das bibliotecas participantes adquirirem mais experiência com as operações.

A descrição realizada no formato MARC beneficia as bibliotecas não apenas no compartilhamento, cooperação, uso e padronização das informações, mas também na organização, preservação e recuperação das informações bem como a atualização dos registros bibliográficos manuais ou *on-line*.

c) Protocolo Z39.50 – criado para promoção e compartilhamento de catálogos em diferentes fontes, permitindo a comunicação para a recuperação das informações desejadas. Para Romani e Borszcz (2006, p. 114), o protocolo Z39.50 é um conjunto de convenções para o intercâmbio de dados, com regras de formato e controle de entrada, transmissão e saída de dados.

O protocolo Z39.50 “[...] permite a consolidação, em redes eletrônicas, de interfaces de busca mais flexíveis, a criação de uma infra-estrutura de informação mais eficiente na identificação e localização de materiais em âmbito mundial” (ROSETTO, 1997). Facilitando assim o processo de comunicação entre diferentes bibliotecas, com intuito de recuperar as informações em cada ambiente.

Do mesmo modo, Cortê (2002, p. 40) afirma que o protocolo Z39.50 serve para acessar catálogos de outras instituições, independentemente do sistema utilizado, promovendo acesso simultâneo aos catálogos, sendo um instrumento de compartilhamento de registros bibliográficos.

d) ISO2709 - desenvolvida pela Library of Congress, com intuito de prover o intercâmbio de dados entre instituições, permitindo a troca de informações e agilizar o processamento técnico das publicações.

Esta norma descreve os requisitos para o formato de intercâmbio de registros bibliográficos determinando todas as formas de documentos sujeitos à descrição bibliográfica. Além disso, não define a extensão do conteúdo de documentos individuais e nem designa significado algum para os parágrafos, indicadores ou identificadores, sendo essas especificações as funções dos formatos de implementação (CÔRTE et al., 2002).

A norma ISO 2709, especifica os requisitos para um formato de troca que irá realizar registros, que descrevem todas as formas de material capaz de realizar descrição bibliográfica, bem como outros tipos de registros.

Dessa forma, esses acordos de cooperação e de padronização bibliográfica, possibilitaram o desenvolvimento da catalogação cooperativa e facilitaram o processamento e a recuperação da informação em uma unidade de informação.

AS TECNOLOGIAS DE INFORMAÇÃO E COMUNICAÇÃO E A CATALOGAÇÃO COOPERATIVA

O uso das novas Tecnologias de Informação e Comunicação (TICs) tem disponibilizado recursos como o computador, a Internet e todas as suas ferramentas a serviço de todos, possibilitando a criação de novas formas de interação entre as pessoas.

O surgimento das TICs trouxe novas perspectivas também para as bibliotecas devido à necessidade de facilitar rapidez no acesso e transferência de informação para os usuários e pelo fato de o mercado exigir que o profissional da informação atualize-se e dinamize os produtos e serviços da Unidade Informacional.

Com o uso das TICs as bibliotecas inovaram produtos e serviços, a noção de valor agregado à informação ganha corpo, as bibliografias foram substituídas por bases de dados, os levantamentos bibliográficos feitos através da cópia xerográfica das fichas catalográficas são realizadas em poucos minutos em catálogos digitais, os boletins ou listas de novas aquisições agora são elaborados com ferramentas do software (sistema) de gerenciamento da biblioteca e disponibilizados pelo próprio sistema, a consulta ao catálogo, livros e periódicos eletrônicos podem ser feitos de qualquer lugar que tenha acesso à internet, suprimindo assim a distância entre a informação e seu usuário (RIBEIRO, 2012).

Diversas formas de implantação e adaptação de produtos e serviços em uma biblioteca são possíveis através das novas tecnologias de informação e comunicação. Os serviços e produtos novos ou melhorados passam uma imagem de inovação e dinamismo das bibliotecas, pois essa imagem é fundamental para conquista e manutenção dos usuários nas bibliotecas.

Como afirma Ribeiro (2012), através da adoção de tecnologia da informação e comunicação as bibliotecas tentam conciliar o padrão da biblioteca de guarda e conservação do acervo para o paradigma atual da biblioteca de apropriação do conhecimento e gerenciamento do acesso à informação.

Com os avanços das tecnologias muitas unidades de informação passaram a oferecer seus serviços de forma virtual com o propósito de facilitar a localização e o acesso aos documentos de forma remota. O uso

das TICs são indispensáveis para proporcionar serviços aos usuários de forma ágil e eficiente em qualquer momento e em qualquer lugar, agiliza o fluxo e a transmissão de informações de forma mais eficiente. Devem-se considerar, também, as questões: flexibilidade e adaptabilidade ao perfil do novo usuário (SOUSA, 2007).

Ainda de acordo com Sousa (2007), a incorporação das tecnologias e suas aplicações, a capacitação dos profissionais da informação, a prestação dos serviços de maneira satisfatória a atender as necessidades dos usuários são pontos que vem suscitando constantes questionamentos e avaliações dos conceitos e métodos estabelecidos aos serviços de referência das bibliotecas universitárias.

Observa-se, portanto, que as tecnologias de informação e comunicação, ao permitir manipulação de diferentes mídias (texto, imagem e som), possibilitam maior interação dos usuários com a biblioteca, pois é um fator de interferência na relação com os usuários. E com isso o emprego dessas tecnologias trouxe maior rapidez no acesso e transferência da informação.

SophiA Biblioteca: estrutura e funcionamento

O SophiA biblioteca é um software de automação denominado de software proprietário, cuja reprodução, redistribuição ou alteração são vedados pelo seu criador ou distribuidor. Esse tipo de software é de acesso pago ou deve-se solicitar autorização para utilizá-lo ao proprietário.

O SophiA biblioteca foi criado em 1993 por profissionais do ITA e da Unicamp, da empresa Prima Informática, que é desde sua criação, especialista no desenvolvimento de soluções tecnológicas para o gerenciamento de instituições de ensino curriculares, escolas de cursos livres, bibliotecas e acervos bibliográficos e não-bibliográficos (PRIMA, 2012).

O SophiA permite a informatização da biblioteca de acordo com as necessidades da instituição sendo cada versão composta pelo módulo gerenciador e por módulos opcionais que permitem ao usuário expandir o potencial do sistema (DANTAS; GOTTSCHALG-DUQUE, 2010).

De acordo com Côte (2002, p. 194) o SophiA tem sua melhor performance no sistema operacional Microsoft Windows, no ambiente de rede Windows e no software de banco de dados MS-SQL Server, além de ser desenvolvido em linguagem Delphi, onde contempla todos os requisitos necessários para o processo de automação de uma biblioteca, destacando-se a manipulação completa de um registro MARC.

O SophiA Biblioteca oferece como vantagens um software robusto, flexível e eficiente; excelente custo-benefício; interface moderna; segurança e controle; flexibilidade nas configurações de circulação; integração com sistemas administrativos (PRIMA, 2012).

DESCRIÇÃO METODOLÓGICA

A concretização desta pesquisa foi fundamentada em algumas teorias científicas no que tange a metodologia do trabalho científico.

Utilizou-se a pesquisa bibliográfica como procedimento inicial de todo e qualquer estudo, na qual foi realizada em diferentes bibliotecas universitárias, bem como em trabalhos disponibilizados em bibliotecas digitais e artigos da Internet, e em bases de dados. Utilizou-se de dados ou de categorias teóricas já trabalhados por outros pesquisadores e devidamente registrados (SEVERINO, 2007, p. 122).

E ainda da pesquisa documental, por se referir aos documentos relacionados às bibliotecas estudadas e ao sistema de automação SophiA Biblioteca, onde a principal característica deste tipo de pesquisa está no fato de que a fonte de coleta de dados está restrita a documentos, escritos ou não, constituindo o que se denomina de fontes primárias (MARCONI; LAKATOS, 2007).

A pesquisa qualitativa priorizou somente as 3 bibliotecárias da empresa. Teve como instrumentos de coleta de dados, um questionário composto por 24 questões, com perguntas abertas e fechadas com o objetivo de analisar os dados referentes à temática abordada, bem como verificar a opinião pessoal dessas profissionais; e ainda da técnica de observação que foi fundamental no processo da pesquisa.

E por fim teve respaldo na pesquisa exploratória, segundo a qual Gil (2010, p. 27) afirma que esta é desenvolvida com o objetivo de proporcionar uma visão geral acerca do objeto estudado e, tendo como estudo de campo as bibliotecas da Eletrobras Eletronorte.

Caracterização do ambiente

As Centrais Elétricas do Norte do Brasil S. A. – Eletronorte, sociedade anônima de economia mista e subsidiária das Centrais Elétricas Brasileiras S. A. – Eletrobras, é uma concessionária de serviço público de energia elétrica. Criada em 20 de junho de 1973, com sede no Distrito Federal, gera e fornece energia elétrica aos nove estados da Amazônia Legal – Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins. Por meio do Sistema Interligado Nacional (SIN), também forne-

ce energia a compradores das demais regiões do País (CENTRAIS ELÉTRICAS DO NORTE DO BRASIL, 2012).

A empresa tem como missão atuar nos mercados de energia de forma integrada, rentável e sustentável. Sua visão é em 2020 ser o maior sistema empresarial global de energia limpa, com rentabilidade comparável as das melhores empresas do setor elétrico. Seus valores englobam foco em resultados, empreendedorismo e inovação, valorização e comprometimento das pessoas, ética e transparência.

As unidades de informação da Eletrobras Eletronorte reúnem importantes acervos que retratam o desenvolvimento do setor elétrico no país. Até o ano de 2008, possuíam realidades totalmente distintas, pois trabalhavam de forma desintegrada e com softwares de automação diferentes. Na sede utilizava-se o sistema Thesaurus, no Maranhão utilizava-se software desenvolvido em SQL e no Pará utilizava-se sistema desenvolvido em Lotus. Assim, a Eletrobras Eletronorte adquiriu o sistema de automação, SophiA biblioteca, com o objetivo de unificar os acervos de suas três unidades de informação.

Oliveira (2009) aponta que antes da integração das bases de dados, o acesso à documentação bibliográfica era realizado por meio do acesso a três catálogos diferentes. Além disso, havia problemas como a não integração entre as bibliotecas, o intercâmbio com outras bibliotecas era realizado manualmente, o procedimento para empréstimo era manual, a reserva de publicações e histórico do usuário era realizada somente no balcão a pedido do usuário, o controle de atrasos na devolução de livros era realizado por meio de listagens de alerta, enfim o serviço era moroso, pouco confiável e ineficaz.

Desse modo, a incorporação do sistema SophiA Biblioteca às bibliotecas da Eletrobras Eletronorte permitiu

A conversão dos dados existentes, distribuídos por três sistemas diferentes para o formato SophiA, num banco único em plataforma SQL. Essa conversão foi levada a efeito pela própria equipe desenvolvedora do SophiA, com assessoria da bibliotecária responsável pelo processamento técnico da Biblioteca, quanto aos formatos e convenções empregados no sistema anterior; a operação em regime de teste, em paralelo com o sistema anterior, quando se detectaram e corrigiram diversos problemas ocorridos na conversão; a interrupção das atividades da Biblioteca por dois dias para a reconversão dos dados atualizados e transição para o novo sistema. (OLIVEIRA, 2009)

Com isso, a aquisição do SophiA Biblioteca possibilitou a migração das obras cadastradas nos antigos sistemas, permitindo que as bibliotecas trabalhassem de forma interligada e em tempo real, na medida em

que cada uma pudesse fazer pesquisas nos catálogos das três unidades e compartilhar informações.

Essa interligação entre as bibliotecas facilitou tanto o ambiente de trabalho dessas unidades quanto do próprio usuário, que pode apreciar os acervos das diferentes regionais sem sair do seu Estado.

a) Biblioteca Monteiro Lobato – Regional Maranhão

No Maranhão, encontra-se a Biblioteca Monteiro Lobato, criada em março de 1997, com o objetivo de suprir as necessidades de informação de seus clientes, compostos de colaboradores e seus dependentes, estagiários e usuários da comunidade, oferecendo a estes, serviços e produtos com qualidade e confiabilidade para satisfação de suas necessidades informacionais. A biblioteca Monteiro Lobato “permite o gerenciamento de toda a informação produzida, tanto a nível interno quanto externo da Eletrobras Eletronorte, provendo os colaboradores informações variadas” (ARAÚJO, 2005).

O ambiente físico da biblioteca é composto por acervo bibliográfico de livros, revistas, relatórios e folhetos; Centro de Estudo Local (CEL), local destinado aos usuários fazerem pesquisas na internet, e videoteca. Os serviços oferecidos pela biblioteca Monteiro Lobato consiste em empréstimo domiciliar, para colaboradores em geral e dependente; reserva on-line através do SophiA biblioteca; consulta local; serviço de referência; acesso à internet; impressão de textos; digitalização de documentos através de scanner; informativo bibliográfico on-line, entre outros.

A biblioteca Monteiro Lobato com objetivo de estimular a leitura dos seus colaboradores e comunidade, já realizou alguns projetos, tais como: o Dia do Livro Infantil, idealizado pela bibliotecária Fátima Lima e realizado durante 10 anos; e o projeto Árvore do Saber, que foi uma parceria da biblioteca Monteiro Lobato com a divisão de Meio Ambiente da empresa. O projeto surgiu em 2007, com o objetivo de despertar a atenção de alunos, comunidade e funcionários para a importância da educação ambiental como um fator determinante para a melhoria do ambiente em que se vive.

b) Biblioteca Raul Garcia Llano – Sede (Brasília)

Segundo Oliveira (2009) a primeira unidade de informação da Eletronorte foi o Centro de Documentação Técnica (CDT), instituída em 1974. Após alguns anos foi criada a Gerência de Suporte à Gestão do Conhecimento (GSEC), inserida na Universidade Corporativa da Eletronorte (UCEL), Superintendência de Desenvolvimento e Educação Empresarial

(GSE), subordinada à Diretoria de Gestão Corporativa (DG). Nessa trajetória, em 2007, o CDT recebeu o nome do primeiro presidente da Eletronorte, passando a chamar-se Biblioteca Raul Garcia Llano (BRGL).

Essa biblioteca tem como objetivo representar o suporte informacional técnico necessário ao desempenho das competências institucionais da Eletronorte, como uma ponte entre o usuário e a informação. Os produtos e serviços oferecidos disponibilizados pela BRGL são: os produtos de Pesquisa e Desenvolvimento (P&D), o acervo on-line, o acesso ao Banco de Monografias, Dissertações e Teses (BMDT), acesso a normas técnicas nacionais e internacionais em formato impresso, acesso à coleção memória técnica (cerca de 5000 publicações provenientes da área de Meio Ambiente e outros acervos em fase de estudos) e intercâmbio com outras bibliotecas para o acesso a artigos técnico-científicos nacionais e internacionais.

c) Centro de Documentação Técnica (CDT) – Regional do Pará

Nessa regional, encontra-se o Centro de Documentação Técnica (CDT), vinculado à Gerência da Regional de Transmissão do Pará. Em 1993, deu início a composição do acervo desta unidade de informação.

Yung e Barbosa (2012, p. 59) relatam que o acervo do CDT inicialmente foi composto por catálogos, instruções técnicas, manuais e planilhas de comissionamento, doados pelos colaboradores. No entanto, conforme crescia a necessidade, o CDT passou a incorporar documentos técnicos relacionados a vários assuntos ligados a Eletrobras Eletronorte.

Esta unidade de informação tem como missão disponibilizar informações que sejam de interesse organizacional. Além disso, o CDT “visa atuar como agente disseminador de informação técnico-científica e contribuir para a aprendizagem contínua e a plena satisfação de seus usuários (YUNG E BARBOSA, 2012, p. 59).

Aliando os negócios da empresa, o CDT tem como objetivos desenvolver-se como agente difusor da informação técnica-científica, contribuir para o desenvolvimento e aperfeiçoamento do corpo técnico e adquirir material de forma sistemática.

CATALOGAÇÃO COOPERATIVA NA ELETROBRÁS ELETRONORTE: discussão e análise dos dados

Para melhor compreensão acerca dos dados coletados, foram agrupados em duas categorias de análise, com vistas a conhecer como ocorre

a aplicabilidade da catalogação cooperativa nas bibliotecas da Eletrobras Eletronorte utilizando o SophiA Biblioteca.

Para manter o anonimato das respondentes, serão utilizadas as letras A, B e C para expor as opiniões das mesmas de modo aleatório em relação a cada temática abordada.

Catalogação Cooperativa: formalização e processamento da informação

Nesta seção a intenção é examinar como as unidades de informação da Eletrobras Eletronorte trabalham a catalogação cooperativa e identificar a estrutura e o funcionamento desse serviço por tais instituições. O quadro 1 mostra esse processo.

Quadro 1 – Catalogação cooperativa: formalização e processamento da informação

Variáveis	Sim			Não		
	Respon- dente A	Respon- dente B	Respon- dente C	Respon- dente A	Respon- dente B	Respon- dente C
Em sua opinião as bibliotecas da Eletrobras Eletronorte aplicam a catalogação cooperativa?	X	X	X			
Houve alguma reunião entre os bibliotecários para que estipulassem os instrumentos de padronização da informação para que a catalogação cooperativa fosse efetivada?				X	X	X
Há uma política de catalogação nessas bibliotecas?				X	X	X
Você acredita que a prática da catalogação cooperativa é importante em suas atividades?	X	X	X			
Quando utiliza a catalogação cooperativa, você verifica se os dados de outra biblioteca estão apropriados?	X	X	X			
Além das bibliotecas da Eletrobras Eletronorte, você recorre a outras unidades de informação utilizar a catalogação de um item?	X	X	X			

Quando há dúvida em catalogar determinado material, você pede ajuda a outro bibliotecário da Eletrobras Eletronorte?	X	X	X			
Há interação entre os bibliotecários nesse processo?	X	X	X			

Fonte: Elaborado pela Autora

Observa-se no quadro 1 que as bibliotecárias utilizam a catalogação cooperativa em suas atividades. Diante disso, “as bibliotecas há muito reconheceram a impossibilidade de isoladamente possuírem todos os recursos necessários para atender as necessidades de seus usuários” (CUNHA, 1987, p. 23 apud ANZOLIN, 2007, p. 496). A automação de bibliotecas e a criação de padrões de intercâmbio de dados facilitaram diferentes bibliotecas realizarem suas atividades de modo cooperativo.

Paralelamente para que haja catalogação cooperativa é necessário que haja padronização nessa atividade. O quadro 1 mostra que não houve reunião para que a catalogação cooperativa fosse efetivada. Para tanto justificaram que:

“Houve uma reunião quando foi implantado o software de bibliotecas (SophiA) e nessa foram discutidos, manuais, metodologias e regras a serem seguidas, mas não especificamente em relação a catalogação cooperativa. A catalogação cooperativa aconteceu naturalmente entre as bibliotecas e com a utilização do novo software.” (Respondente A)

“Não houve uma reunião prévia sobre esse assunto, mas quando da implantação do software SophiA a empresa ministrou treinamento aos profissionais da área, orientando-os sobre as metodologias a serem utilizadas nas bibliotecas da Eletrobras Eletronorte.” (Respondente B)

“Houve reunião para implantação do SophiA.” (Respondente C)

Entretanto, essas bibliotecas passaram a realizar suas atividades de acordo com algumas metodologias estipuladas quando da implantação do sistema de automação SophiA Biblioteca, mediante o manual de processamento técnico dessas unidades de informação, subtendo-se a utilização de um padrão ao catalogar um documento.

Santos (2008, p. 19) considera que se torna essencial a existência de padrões que possibilitem uma interpretação uniforme e universal, em qualquer idioma e em qualquer tipo de unidade de informação, por catalogadores e usuários nos mais diversos ambientes informacionais.

Na mesma direção da questão anterior, o quadro 1 aborda uma política de catalogação. Desse modo, as respondentes afirmam não existir uma política, porém destacaram a existência de um manual de processamento técnico nas bibliotecas que estabelece critérios gerais para o pro-

cessamento técnico de documentos, em qualquer suporte físico, desde sua aquisição até descarte.

Segundo Silva (2009, p. 1) uma política de catalogação é essencial para a constituição e funcionamento de redes de bibliotecas à escala local, regional, nacional e internacional, e para o fornecimento de serviços biblioteconômicos de qualidade, num mundo em que, de modo complexo e não linear, as pressões intensas da globalização afirmam-se em paralelo com uma crescente afirmação das particularidades e valores locais e regionais.

Outro dado expresso no quadro 1 é sobre a importância da prática da catalogação cooperativa em suas atividades. Nesse sentido, as respondentes ressaltaram que:

“A grande vantagem que a cooperação traz é a diminuição do tempo e a redução dos custos.” (Respondente A)

“Diminui o tempo ao tratar a informação.” (Respondente B)

“Agiliza o trabalho do bibliotecário e há uma maior padronização da informação.” (Respondente C)

Outro dado observado no quadro 1 é em relação à verificação dos dados de outra biblioteca quando utilizada a catalogação cooperativa. As respondentes destacam que:

“A catalogação cooperativa não é só copiar e colar. É necessário o controle de qualidade dos registros.” (Respondente A)

“Os dados obtidos são analisados a fim de mantermos o padrão da biblioteca.” (Respondente B)

Perguntadas sobre a utilização da catalogação cooperativa além das bibliotecas da Eletrobras Eletronorte:

Quanto mais unidades e catálogos forem utilizados maior será a possibilidade de catalogação cooperativa. (Respondente A)

Nem sempre os dados da catalogação são encontrados nas bibliotecas da Eletrobras Eletronorte. (Respondente B)

Quanto mais bibliotecas consultadas, mais facilidades. (Respondente C)

O quadro 1 também identifica a questão sobre pedir ajuda de outro bibliotecário quando houver uma dúvida ao catalogar algum material. Foi observado que as respondentes afirmam positivamente esse questionamento, onde:

“A troca de informações é muito importante e enriquecedora.” (Respondente A)

“A troca de informações auxilia no processo de catalogação”. (Respondente B)

“Para evitar erros e manter o padrão de qualidade da catalogação”. (Respondente C)

Ainda de acordo com o quadro 1, as bibliotecárias afirmam interajam no processo de catalogação cooperativa:

Para manter a qualidade e a padronização catalogação. (Respondente A)

A interação é fundamental. (Respondente B)

A interação entre os bibliotecários é fundamental para o andamento do trabalho. (Respondente C)

Conforme Santos (2008, p. 163) a cooperação exige interação, colaboração, objetivos comuns, atividades e ações conjuntas e coordenadas, relações de respeito mútuo, uma postura de tolerância e convivência com as diferenças e um processo de negociação constante. Acrescenta-se ainda que no ambiente cooperativo, o que se espera é aumentar o potencial do grupo, fazendo com que o resultado seja alcançado, de forma maior que a soma das contribuições individuais de cada membro do grupo.

As Tecnologias da Informação e Comunicação e a Catalogação Cooperativa

Esta seção refere-se aos aspectos relacionados às Tecnologias da Informação e Comunicação, envolvendo o sistema de automação SophiA Biblioteca e a catalogação cooperativa, conforme os dados mostrados no quadro 2.

Quadro 6 – As Tecnologias da Informação e Comunicação e a Catalogação Cooperativa

Variáveis	Sim			Não		
	Respondente A	Respondente B	Respondente C	Respondente A	Respondente B	Respondente C
Em relação ao sistema de automação SophiA Biblioteca, este preenche as necessidades da biblioteca?	X	X	X			
O SophiA Biblioteca atende suas exigências enquanto bibliotecário?	X	X	X			
Com a adoção do SophiA Biblioteca, houve impacto nas rotinas da biblioteca?	X	X	X			

O sistema é de fácil utilização na tarefa de catalogação?	X	X	X			
A aplicabilidade da catalogação cooperativa no SophiA biblioteca permite a diminuição do tempo no tratamento da informação?	X	X	X			

Fonte: Elaborado pela Autora

Conforme o quadro 2, percebe-se que no primeiro item as bibliotecárias acreditam que o SophiA Biblioteca preenche as necessidades da unidade de informação. Elas relatam que:

“A biblioteca teve um crescimento significativo do seu acervo tratado tecnicamente. Em virtude das características do SophiA Biblioteca Avançado, a equipe da Biblioteca realiza um trabalho ágil e de qualidade no atendimento às necessidades informacionais dos colaboradores da Eletronorte.” (Respondente A)

“Com a implementação do SophiA Biblioteca houve uma otimização nos processos da biblioteca.” (Respondente B)

Um fato determinante é que o software da unidade de informação além de atender as necessidades de informação, deve ser “compatível com o desenho e a cultura organizacional, com o parque computacional instalado, tamanho do acervo e o perfil de seus usuários [...]” (CÔRTE ET AL., 2002, p. 55).

Em relação ao SophiA Biblioteca atender as exigências do bibliotecário, o quadro 2 mostra que as respondentes acreditam que sim, pois isso se evidencia nas seguintes falas:

“A utilização do SophiA permite, entre outros benefícios, a catalogação cooperativa de materiais bibliográficos, de acordo com padrões internacionais de descrição bibliográfica e intercâmbio de dados; a automatização de forma integrada de procedimentos de aquisição, processamento técnico, circulação, controle, empréstimo e disseminação de recursos informacionais; aperfeiçoamento dos mecanismos de recuperação de informações; otimização da utilização dos recursos informacionais, facilitando a solicitação e efetivação de empréstimos, renovações e reservas; o desenvolvimento de novos produtos e serviços de informação para melhor atender às necessidades dos clientes”. (Respondente A)

“O SophiA é um programa de interface fácil e um facilitador no processo de catalogação, geração de relatórios, estatística, circulação, etc.” (Respondente B)

Côrte et al. (2002, p. 25) analisa que as bibliotecas “[...] devem utilizar um sistema automatizado que privilegie todas as etapas do ciclo do-

cumental, onde a escolha do software recaia sobre uma ferramenta que contemple os recursos hoje disponíveis”.

Outro dado demonstrado no quadro 2 questiona se houve impacto nas rotinas dessas bibliotecas. As respondentes justificam da seguinte forma:

“Houve a integração entre as bibliotecas”. (Respondente A)

“Um impacto significativo foi a disponibilização do SophiA WEB (Intranet) para os empregados da Eletrobras Eletronorte e a integração entre as bibliotecas.” (Respondente B)

“Entre as melhorias que podem ser identificadas após a aquisição do SophiA pode-se destacar a integração das bases de dados dos acervos das unidades de biblioteca da Eletronorte, reunindo informações da biblioteca da sede, em Brasília, de Belém/PA e de São Luís/MA, representando um significativo avanço à gestão de bibliotecas desta Empresa. Além disto, destacamos: a padronização dos registros de acordo com as normas existentes; a velocidade para processamento técnico de materiais do acervo; a facilidade, agilidade e confiabilidade para as atividades relacionadas à Seção de Referência. A implementação do SophiA trouxe a autonomia no uso do sistema com consulta integrada; o intercâmbio com outras bibliotecas é realizado com a importação e exportação de dados, no padrão MARC21, utilizando protocolo Z39.50; a reserva de publicações e histórico do usuário é realizada pelo próprio usuário através do uso de senha; o controle de atrasos é realizado por meio de lembretes automatizados e personalizados via e-mail.” (Respondente C)

No que tange esses impactos, Figueiredo (1998 apud CÔRTE et al., 1999, p. 242) identifica que o maior benefício da informatização é a rapidez, agilidade e eficiência no atendimento e prestação de serviços, isto é, a otimização das atividades não só com relação aos usuários, como também no que diz respeito ao controle e formação do acervo, levantamentos bibliográficos, catalogação, empréstimos, comutação, reclamação de obras em atraso e processamento técnico.

Ainda no quadro 2 é verificado se o SophiA Biblioteca é de fácil utilização na tarefa de catalogação, sendo que as respondentes afirmam positivamente.

“Apenas como exemplo na Biblioteca X, no ano de 2008, contratou-se 8 estagiários de biblioteconomia pelo período de 6 meses para o processamento técnico no SophiA de determinada documentação bibliográfica existente. A meta daquele trabalho foi o tratamento técnico de 8000 documentos do acervo da Biblioteca X. O número total de documentos processados foi 11.450, atingindo e superando a meta inicial desta atividade”. (Respondente A)

“Porque permite a utilização de catalogação cooperativa e utilização de catálogos do próprio SophiA”. (Respondente B)

Côrte et al. (1999, p. 248) enfatiza que os modernos softwares de automação de bibliotecas estão facilitando esta tarefa, deixando a cargo do sistema a transformação dos dados de catalogação para o formato MARC, obedecidas, evidentemente, as características mínimas de descrição bibliográfica.

De acordo com o quadro 2 é verificado se a aplicabilidade da catalogação cooperativa utilizando o SophiA Biblioteca permite a diminuição do tempo do bibliotecário ao tratar uma informação. Visualizamos que todas as respondentes abordam que sim.

“Porque quando uma obra já está catalogada em uma das bibliotecas da Eletrobras Eletronorte basta acrescentar os dados do exemplar. Além disso, o SophiA permite a duplicação da obra para os casos de mudança de edição”. (Respondente A)

“O livro é catalogado apenas uma vez.” (Respondente B)

“A partir da implantação do SophiA o livro é catalogado apenas um vez. Assim as outras bibliotecas da rede colocam apenas a informação do exemplar. Além da possibilidade de importação de registros do Portal SophiA, da Rede RVBI e do Z39.50, entre outros.” (Respondente C)

E por fim, foi pedido para as respondentes citarem pelo menos um aspecto positivo e/ou negativo, do SophiA Biblioteca no que tange ao trabalho de catalogação. Assim, ressaltaram que:

“A integração e o compartilhamento dos recursos informacionais de cada uma das unidades, além da sistematização e padronização do processamento técnico dos acervos, otimizando a gestão e o intercâmbio de informações entre a Biblioteca da Sede e a das Unidades Descentralizadas facilitam o acesso à informação pelos usuários. Com o SophiA, as Bibliotecas da Eletronorte trabalham em ambiente único”. (Respondente A)

“Sistematização e padronização do processamento técnico”. (Respondente B)

“Com a utilização do SophiA a publicação é catalogada apenas uma vez (exceto quando muda de edição mas, nesse caso, pode-se duplicá-lo e atualizar os dados de edição); padronização do processamento técnico.” (Respondente C)

CONCLUSÃO

Com os avanços da Tecnologia da Informação e Comunicação muitas atividades moldaram-se. A catalogação cooperativa é uma delas, ao passo que, esta era feita manualmente através de fichas impressas e atu-

almente é feita em tempo real em questão de minutos, através de uma máquina que permite a utilização de normas e protocolos de comunicação interoperante, por meio da Internet.

No decorrer deste estudo, houve a preocupação em oferecer elementos que demonstrassem claramente como ocorre a catalogação cooperativa em uma unidade de informação.

Tendo em vista o objetivo geral que consistiu em fazer uma análise da aplicabilidade da catalogação cooperativa nas bibliotecas da Eletrobras Eletronorte, através do sistema de automação SophiA Biblioteca; e os objetivos específicos demonstrando como é realizada a catalogação cooperativa nessas bibliotecas; identificando os instrumentos de padronização internacional utilizados na catalogação cooperativa e no sistema SophiA; e verificando ainda se o sistema preenche o perfil dessas bibliotecas, pôde-se observar que a catalogação cooperativa utilizando o sistema de automação SophiA Biblioteca exerce grande importância no ambiente informacional estudado, pois a primeira constatação estudando as unidades de informação da Eletrobras Eletronorte é que o sistema de automação permite a interligação de atividades desses locais e a segunda, é que os bibliotecários responsáveis aplicam a catalogação cooperativa ao realizar a representação descritiva.

Os dados obtidos através do questionário possibilitou um maior entendimento em relação à catalogação cooperativa sabendo-se como ocorre esse trabalho, quais os instrumentos que permitem que seja efetiva e quais eram os impactos dessa atividade nas unidades de informação estudadas.

De fato, foram verificados vários pontos positivos da catalogação cooperativa nessas unidades de informação, utilizando o SophiA Biblioteca. Dentre esses fatores, vale ressaltar a importância do trabalho cooperativo, principalmente quando o resultado ou o produto é levar informação a outras pessoas. No nosso ambiente de estudo, a integração dos profissionais e a interligação do sistema ao utilizar-se da catalogação cooperativa tornou-se essencial essa atividade.

Outro fato relevante está no SophiA Biblioteca, compreendido nesta pesquisa como um sistema de automação de fácil utilização, que atende as necessidades da biblioteca e do bibliotecário e que permite que as bibliotecas da mesma instituição e de diferentes localidades do Brasil, no Maranhão, Pará e Brasília, possam trabalhar de modo integrado, compartilhando suas informações.

A revisão de literatura foi sem dúvida nenhuma essencial para dar consistência ao trabalho proposto e por reforçar a compreensão de conceitos e importância da catalogação cooperativa, as Tecnologias de Informação e Comunicação envolvidas nesse processo, e os demais aspectos abordados neste estudo.

Dentre as contribuições que este trabalho pode ocasionar, pode-se destacar a relevância para uma prática efetiva em relação ao uso da catalogação cooperativa, pois está aliada ao campo da TIC, ocorrendo de forma padronizada, permite um menor tempo ao tratar a informação, além de que expande o acervo informacional além dos muros físicos da unidade de informação. Daí a importância do bibliotecário participar de cursos de treinamento e aperfeiçoamento profissional, sendo aberto às mudanças e pronto para ajudar e/ou sugerir melhorias no serviço do outro.

Outra contribuição refere-se aos estudantes do curso de Biblioteconomia, na medida em que adentrando ao mercado de trabalho, está preparado para trabalhar de modo cooperativo, seja catalogando, indexando, emprestando materiais, entre outras funcionalidades.

REFERÊNCIAS

ANZOLIN, Heloísa Helena. Atualizações em AACR2. In: ENCONTRO NACIONAL DOS USUÁRIOS DA REDE PERGAMUM. 9, 2007, Curitiba. **Anais...** Curitiba: Pontifícia Universidade Católica do Paraná – Sistema Integrado de Bibliotecas (SIBI), 2007. Disponível em: <http://cdij.pgr.mpf.gov.br/sistema-pergamum/ix-encontro-nacional/20_04_2007/Curso%20AACR2.pdf>. Acesso em: 01 dez. 2012.

ARAÚJO, Lindanira. **Projeto: Divulgando a biblioteca da CMA**. São Luís, 2005

BARBOSA, Alice Príncipe. **Novos rumos da catalogação**. Rio de Janeiro: BNG / Brasilart, 1978. 246 p.

CAMPELLO, Bernadete. **Introdução ao controle bibliográfico**. 2. ed. Brasília, DF: Briquet de Lemos, 2006. 94 p.

CENTRAIS ELÉTRICAS DO NORTE DO BRASIL. **Eletrobras Eletronorte**. Disponível em: <<http://www.eln.gov.br/opencms/opencms/>>. Acesso em: 14 maio 2012.

CÔRTE, Adelaide Ramos et al. Automação de bibliotecas e centros de documentação: o processo de avaliação e seleção de softwares. **Ci. Inf.**, Brasília, v. 28, n. 3, p. 241-256, set./dez. 1999. Disponível em: <<http://www.scielo.br/pdf/ci/v28n3/v28n3a2.pdf>> . Acesso em: 08 set. 2012.

CÔRTE, Adelaide Ramos et al. **Avaliação de softwares para bibliotecas e arquivos**. 2. ed. rev. e ampl. São Paulo: Polis, 2002.

DANTAS, Suellen Fernande; GOOTTSCHALG-DUQUE, Cláudio. Usabilidade: um estudo de caso dos WEBOPACS dos sistemas Aleph e SophiA na biblioteca Ministro Rubem Rosa do TCU, **Biblionline**, João Pessoa, v. 6, n. 1, p. 10-24, 2010. Disponível em: <<http://periodicos.ufpb.br/ojs/index.php/biblio/article/view/4901>>. Acesso em: 30 abr. 2012.

GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2010. 200 p.

GRAVINA, Roberta Amaral Sertório. A catalogação cooperativa no Brasil e o modelo espanhol: um estudo comparado entre a rede Bibliodata e a rede de bibliotecas do Instituto Cervantes. In: SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA FESPSP, 2, São Paulo. **Anais...** São Paulo: FESP, 2012. Disponível em: <http://www.fespsp.org.br/sic/papers/2010_2/IISIC_Catalogacao_Cooperativa.pdf>. Acesso em: 21 nov. 2012.

LOPES, Marcos Luiz Pereira. **Catalogação cooperativa em redes de informação**: estudo de caso da rede Bibliodata. 2010. Trabalho de Conclusão de Curso (Graduação)-Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2010.

MARCONI, Maria de Andrade; LAKATOS, Eva Maria. **Fundamentos de metodologia científica**. 6. ed. São Paulo: Atlas, 2007.

MEY, Eliane Serrão Alves; SILVEIRA, Naira Christofolletti. **Catalogação no plural**. Brasília: Briquet de Lemos, 2009. 217 p.

OLIVEIRA, Jeanne Lucena. **Acesso à informação na Eletronorte após a implantação do SophiA (Sistema de Automação de Bibliotecas)**. [Brasília, DF: Eletronorte], 2009.

ORTEGA, Cristina Dotta. **Informática documentária**: estado da arte. 2002. 259 f. Dissertação (Mestrado em Ciência da Comunicação)- Escola de Comunicação em Artes da Universidade de São Paulo, São Paulo, 2002.

PRIMA informática: softwares para gestão de escolas e bibliotecas. **SophiA Biblioteca**. Disponível em: <<http://www.primasoft.com.br/2006/html/institucional.php>>. Acesso em: 12 jun. 2012.

RIBEIRO, Rejane M. R. A tecnologia da informação e comunicação (tic): fator condicionante da inovação em bibliotecas universitárias, **Rev. Dig. Bibl. Ci. Inf.** Campinas, v.9, n. 2, p. 41-48, jan./jun. 2012.

ROMANI, Cláudia; BORSZCZ, Iraci (Org.). **Unidades de informação: conceitos e competências**. Florianópolis: Ed. da UFSC, 2006.

SANTOS, Erika Alves dos. **Catálogo Cooperativa: propósitos, vantagens e desvantagens**. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO (CBBDD), 21, 2005, Curitiba. **Anais...** Curitiba: FEBAB, 2005. Disponível em: <<http://bancodedadosfurg.files.wordpress.com/2010/08/bd-catalogacao-cooperativa.pdf>>. Acesso em: 21 nov. 2012.

SANTOS, Plácida L. V. Amorim da Costa. **Redes informacionais como ambiente colaborativo e de empoderamento: a catalogação em foco**. In: GUIMARÃES, José Augusto Chaves; FUJITA, Mariângela Spotti Lopes (Org.). **Ensino e pesquisa em Biblioteconomia no Brasil: a emergência de um novo olhar**. São Paulo: Cultura Acadêmica, 2008. p. 155-171.

SEVERINO, Antônio Joaquim. **Metodologia do trabalho científico**. 23. ed. São Paulo: Cortez, 2007. 304 p.

SILVA, Armando Jorge Silva. **Política de catalogação para as bibliotecas da Rede Nacional de Bibliotecas Públicas**. Lisboa: Ministério da Cultura, 2009. Disponível em: <http://www.dglb.pt/sites/DGLB/Portugues/bibliotecasPublicas/documentacaoBibliotecas/Documents/PoliticaCatalogacaoDGLBfev2009.pdf>. Acesso em: 29 nov. 2012.

SILVEIRA, Naira Christofolletti; TÁLAMO, Maria de Fátima Gonçalves Moreira. **Os FRBR e a escolha do ponto de acesso pessoal. Perspectivas em Ciência da Informação**, v. 14, n. 2, p. 108-120, maio/ago. 2009.

SOUSA, Beatriz Alves de. **Uso de tecnologias da informação e comunicação (TICs) nos serviços de referência das bibliotecas dos centros federais de educação tecnológica (CEFETs) das regiões norte e nordeste do país**. In: Congresso de Pesquisa e Inovação da Rede Norte Nordeste de Educação Tecnológica, 2, 2007, João Pessoa. **Anais...** João Pessoa: CEFET/PB, 2007.

VASCONCELLOS, Paulo de Avellar de Góes e. **Bibliodata/CALCO - Informação Bibliográfica para o Desenvolvimento. Ciência da Informação**, v. 25, n. 3, 1996.

VETTER, Silvana Maria de Jesus; ARAUJO, Leonardo Pinto. **Padrão MARC21 e catalogação em bibliotecas universitárias de São Luís/MA**. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 17, 2012, Gramado, RS. **Anais eletrônicos...** Disponível em: <www.eventweb.com.br/snbu2012/specific-files/.../index.php?file>. Acesso em: 10 nov. 2012.

YUNG, Daiane da Silva; BARBOSA, Janaína Soares Lopes. **A importância da utilização e atualização de linguagens documentárias em sistemas de informação especializados**. 2012. 107 f. Trabalho de Conclusão de Curso (Graduação)-Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2012. Disponível em: <http://bdm.bce.unb.br/bitstream/10483/3641/1/2012_DaianeYung_Jana%C3%ADnaBarbosa.pdf>. Acesso em: 26 set. 2012.

