UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS SOCIAIS APLICADAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
MESTRADO EM CIÊNCIA DA INFORMAÇÃO

JACIARA CRISTINA ALMEIDA DO AMARAL

USO DE LINGUAGENS DOCUMENTÁRIAS NO CONTEXTO DE REPOSITÓRIOS INSTITUCIONAIS DE UNIVERSIDADES FEDERAIS DA REGIÃO NORTE DO BRASIL

Belém - PA
2020

JACIARA CRISTINA ALMEIDA DO AMARAL

USO DE LINGUAGENS DOCUMENTÁRIAS NO CONTEXTO DE REPOSITÓRIOS INSTITUCIONAIS DE UNIVERSIDADES FEDERAIS DA REGIÃO NORTE DO BRASIL

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação, do Instituto de Ciências Sociais Aplicadas, da Universidade Federal do Pará como parte dos requisitos necessários para obtenção do título de Mestre em Ciência da Informação.

Linha de pesquisa: Organização da Informação
Orientadora: Prof. Dra. Franciele Marques Redigolo
Co-orientador: Thiago Henrique Bragato Barros

Belém - PA

2020

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

A485u Amaral, Jaciara Cristina Almeida do
Uso de linguagens documentárias no contexto de repositórios institucionais de universidades federais da Região Norte do Brasil / Jaciara Cristina Almeida do Amaral. — 2020.
110 f. : il. color.

Orientador(a): Profª. Dra. Franciele Marques Redigolo
Coorientador(a): Prof. Dr. Thiago Henrique Bragato Barros
Dissertação (Mestrado) - Programa de Pós-Graduação em Ciência da Informação, Instituto de Ciências Sociais Aplicadas, Universidade Federal do Pará, Belém, 2020.

1. Indexação. 2. Representação do conhecimento (Teoria da informação). 3. Repositórios institucionais - Brasil, Norte. I. Título.

CDD 025.47

JACIARA CRISTINA ALMEIDA DO AMARAL

USO DE LINGUAGENS DOCUMENTÁRIAS NO CONTEXTO DE REPOSITÓRIOS INSTITUCIONAIS DE UNIVERSIDADES FEDERAIS DA REGIÃO NORTE DO BRASIL

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Pará, como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Informação

Linha de pesquisa: Organização da Informação.

DATA DA AVALIAÇÃO: _____/_____/_____

CONCEITO: _____

BANCA EXAMINADORA

_____

Profª. Drª. Franciele Marques Redigolo

(Orientadora – PPGCI/ICSA/UFPA)

_____

Profª. Drª. Marise Teles Condurú

(Membro – PPGCI/ICSA/UFPA)

_____

Bibliotecária Drª. Milena Polsinelli Rubi

(Membro – UFSCAR)

Dedico este trabalho a Maria Egídia Neves de Almeida,
minha mãe e maior amor.

AGRADECIMENTOS

Agradeço à força criadora da vida (Deus, Jah, Alá, Olorum, Javé, Tupã) que tem me levado por caminhos iluminados de conquistas e aprendizados.

Agradeço aos meus pais, Maria Egídia e Cristino Amaral (*in memorian*), que, mesmo diante das dificuldades da vida, nunca deixaram de me incentivar nos caminhos da educação e do conhecimento.

Aos meus irmãos, grandes inspirações.

À minha querida orientadora, Prof. Dra. Franciele Redigolo, que, como uma verdadeira guia, me acompanhou e conduziu na realização desta pesquisa.

A todos os professores do PPGCI e aos colaboradores do programa – foram momentos de muito aprendizado.

Às membras da banca de qualificação/defesa, Prof. Dra. Marise Telles Condurú e Bibliotecária Dra. Milena Polsinelli Rubi, muito obrigada. O olhar preciso de pesquisadoras, as sugestões e considerações foram muito valiosas para a lapidação desta pesquisa. Muito obrigada pelo aprendizado.

Aos meus colegas de Mestrado: Mateus, Letícia, Cris, Vanessa, Eduardo, Rodrigo, Nicole, Ester, Nara e Glenda, agradeço o companheirismo, apoio e incentivo nas horas mais difíceis. Agradeço especialmente à Layane Gaia, que não me deixou desistir antes mesmo da seleção e foi uma grande amiga e incentivadora durante esta caminhada. Nilzete Gomes, minha "irmã" de pesquisa. Ana Miranda, Anezia Eugênia, Zilah Edelburga, amigas inspiradoras.

Camila de Brito, Denize Escalabrin, Luele Mayara e Valdenise César, obrigada pelo apoio e ajuda com as traduções, correções ortográficas e normalizações.

Allan Rocha, obrigada pelas dicas e ajuda com o Excel.

Aos colegas do Instituto de Ciências Biológicas (ICB/UFPA), ao nome de nosso diretor Ricardo Vieira, obrigada pela compreensão e apoio na conciliação entre o trabalho e o estudo.

Estefany Bonifácio (*Baixinea*), obrigada por me aguentar e sempre acreditar em mim, me incentivando a seguir em frente. Você é incrível!

Às Garotas do Bueiro (não posso revelar suas identidades reais), obrigada pelo apoio, pelos momentos de diversão e descontração! Se alguém não entender, depois eu explico...

“Ninguém caminha sem aprender a caminhar, sem aprender a fazer o caminho caminhando, refazendo e retocando o sonho pelo qual se pôs a caminhar.”
(Paulo Freire)

“Busquem conhecimento!”

(Bilu)

RESUMO

Estudo sobre o uso das linguagens documentárias no contexto de repositórios institucionais universitários da Região Norte do Brasil. Objetiva, especificamente, conceituar Organização do Conhecimento e os caminhos que conduzem à elaboração de linguagens documentárias, assim como as características e principais tipos destas. Traçar um panorama sobre o movimento Acesso Livre, a crise dos periódicos tradicionais e a busca por mais liberdade no acesso a textos científicos, contexto sob o qual desenvolveram-se os repositórios. Demonstrar a adesão ao movimento acesso livre e as primeiras ações em seu favor ao redor do mundo e no Brasil, como as ações de incentivo de pesquisadores e órgãos governamentais. Tem abordagem qualitativa, de caráter explicativo, caracterizando estudo de caso sobre a utilização de linguagens documentárias na padronização de termos representativos de documentos disponibilizados em repositórios institucionais universitários. O instrumento de coleta de dados configurou-se em um questionário semiaberto online, contendo 23 questões, aplicado aos sujeitos da pesquisa (bibliotecários), sendo um respondente de cada repositório. As 23 questões estão divididas em três macros sessões de análise, cada uma tendo por objetivo observar um aspecto norteador da investigação: a) ambientação do repositório e das atividades do bibliotecário, b) verificar a existência de política de indexação e o uso de linguagens documentárias; c) observar os aspectos do método de tradução conceitual adotada pelos indexadores. Foram enviados 8 questionários, sendo um por repositório, e obteve-se 100% (cem por cento) de respostas. A análise dos dados foi feita a partir destas resposta, de acordo com as categorias que compõem as macros sessões de análise, sendo uma proporção de 8/1, ou seja, para cada pergunta foram analisadas 8 respostas, o que possibilitou traçar um perfil da atividade de indexação de documentos nos repositórios digitais universitários federais da Região Norte. Neste sentido, observou-se que o uso das linguagens documentárias nos repositórios pesquisados ainda não é uma unanimidade, pois a ausência de uma política de indexação nas bibliotecas possibilita a livre escolha pela utilização ou não de uma linguagem controlada, prevalecendo ainda o uso da linguagem natural ou o uso desta juntamente com as LDs. Outro fator que reflete na adoção de uma linguagem documentária é o caráter particular da região Amazônica, onde variam os termos, conceitos e grafias, os quais, comumente, não são contemplados pelas linguagens normalmente adotadas, pois estas são mais gerais e não específicas para atender a demanda dos profissionais que trabalham com a indexação de documentos gerados por pesquisas na região. Concluiu-se assim que, apesar da prevalência da linguagem natural, os indexadores têm compreensão sobre a importância do controle terminológico e dos processos que o envolvem, como a leitura técnica e a adequação de remissivas. Falta, porém, as bibliotecas adotem uma política norteadora sobre a indexação, que ajude os profissionais a manterem um catálogo padronizado, recuperável de maneira exaustiva, seja ele de acervo físico ou digital.

Palavras-chave: Indexação. Representação do conhecimento (Teoria da informação). Repositórios institucionais. Brasil, Norte.

ABSTRACT

Study on the use of documentary languages in the context of university institutional repositories in the Northern Region of Brazil. It specifically aims to conceptualize Knowledge Organization and the paths that lead to the elaboration of documentary languages, as well as their characteristics and main types. Draw an overview of the Free Access movement, the crisis of traditional journals and the search for more freedom in accessing scientific texts, the context under which the repositories were developed. Demonstrate adherence to the free access movement and the first actions in its favor around the world and in Brazil, such as incentive actions by researchers and government agencies. It has a qualitative, explanatory approach, featuring a case study on the use of documentary languages in the standardization of representative terms of documents made available in university institutional repositories. The data collection instrument was configured in a semi-open online questionnaire, containing 23 questions, applied to the research subjects (librarians), one respondent from each repository. The 23 questions are divided into three macro analysis sessions, each aiming to observe a guiding aspect of the investigation: a) setting of the repository and the activities of the librarian, b) verifying the existence of an indexing policy and the use of documentary languages ; c) observe the aspects of the conceptual translation method adopted by the indexers. Eight questionnaires were sent, one per repository, and 100% (one hundred percent) responses were obtained. The analysis of the data was made from these answers, according to the categories that make up the macro analysis sessions, being a proportion of 8/1, that is, for each question 8 answers were analyzed, which made it possible to draw a profile of the document indexing activity in federal university digital repositories in the Northern Region. In this sense, it was observed that the use of documentary languages in the researched repositories is not yet unanimous, since the absence of a policy of indexing in libraries allows the free choice for the use or not of a controlled language, still prevailing the use of language natural or the use of it together with LDs. Another factor that reflects in the adoption of a documentary language is the particular character of the Amazon region, where the terms, concepts and spellings vary, which, commonly, are not contemplated by the languages normally adopted, as these are more general and not specific to attend the demand of professionals who work with the indexing of documents generated by research in the region. It was concluded that, despite the prevalence of natural language, the indexers have an understanding of the importance of terminological control and the processes that involve it, such as technical reading and the adequacy of cross-references. However, there is still a lack of libraries to adopt a guiding policy on indexing, which helps professionals to maintain a standardized catalog, which can be recovered in an exhaustive manner, be it from physical or digital collections.

Keywords: Indexing. Knowledge representation (Information theory). Institutional repositories. Brazil, North.

LISTA DE QUADROS

Quadro 1 - Relação dos objetivos com os capítulos da pesquisa 22
Quadro 2 - Características de relacionamento entre os termos nas linguagens documentárias 34
Quadro 3 - Universo da pesquisa 53
Quadro 4 - Distribuição das perguntas e aspectos a serem observados 55
Quadro 5 - Base conceitual para a elaboração das categorias de análise dos dados 56
Quadro 6 - Forma que os dados foram analisados 57
Quadro 7 – ano de início das atividades em cada repositório 60
Quadro 8 - Atividades desenvolvidas pelos indexadores. 62
Quadro 9 - Identificação das obtidas na questão 4 63
Quadro 10 - Especificidade das áreas do conhecimento onde atuam os bibliotecários. 64
Quadro 11 - Existência da política de indexação nas bibliotecas mantenedoras dos repositórios. 66
Quadro 12 - Tipo de linguagem adotada no repositório 68
Quadro 13 – Estratégia do uso de linguagens de indexação nos repositórios 70
Quadro 14 - Indexação de documentos digitais 72
Quadro 15 - Média de termos utilizados pelos indexadores. 73
Quadro 16 - Análise de assunto. 74
Quadro 17 - Execução da análise documentária por parte dos bibliotecários 76
Quadro 18 - Partes do documento utilizadas pelos bibliotecários para análise documentária 78
Quadro 19 – Importância das linguagens documentárias 80
Quadro 20 - Aspectos importantes do uso das linguagens documentárias na tradução conceitual. 81
Quadro 21 - Aspectos da utilização de linguagens documentárias: organização e recuperabilidade 81
Quadro 22 - Especificidade nos reposiórios 83
Quadro 23 - Uso de termos qualificadores e subdivisões 84
Quadro 24 - Orientações sobre a exaustividade nos repositórios. 85
Quadro 25 - Uso de remissivas nos repositórios. 86
Quadro 26 - Estratégias utilizadas pelos bibliotecários em relação às remissivas nos repositórios. 87
Quadro 27 - Tradução dos termos para outros idiomas. 88
Quadro 28 - Resistência dos autores em relação à tradução conceitual. 89
Quadro 29 - Dificuldades enfrentadas na indexação de documentos científicos na Amazônia. 91
Quadro 30 - Atualização das LDs em relação aos termos científicos na Amazônia. 92
Quadro 31 - Relação das categorias de análise e dos resultados obtidos. 95

LISTA DE FIGURAS

Figura 1 - Hierarquização entre repositórios 39
Figura 2 - Instituições que receberam o kit tecnológico para implantação de repositórios institucionais em 2009. 45
Figura 3 - Estrutura hierárquica dentro dos repositórios institucionais 47
Figura 4 - Etapas de submissão de um item dentro de um repositório. 47
Figura 5 - Campo assunto de um item submetido a um repositório. 51
Figura 6 - Tela de navegação pelo índice alfabético de assuntos em um repositório institucional universitário. 52

LISTA DE GRÁFICOS

Gráfico 1 - Respondentes por estado 59
Gráfico 2 - Ano em que cada repositório entrou em funcionamento. 60
Gráfico 3 - Áreas do conhecimento onde atuam os bibliotecários. 63
Gráfico 4 - Atuação dos bibliotecários por área do conhecimento 64
Gráfico 5 - Quantitativo das universidades federais nortistas que possuem política de indexação 66
Gráfico 6 - Tipo de linguagem de indexação utilizada nos repositórios. 69
Gráfico 7 - Análise documentária por parte dos bibliotecários 75
Gráfico 8 - – Partes do texto recomendadas pela teoria e partes utilizadas pelos bibliotecários para análise documentária. 77
Gráfico 9 - Aspectos da utlização das linguagens documentárias. 78
Gráfico 10 - Especificidade nos repositórios. 82
Gráfico 11 - Especificidade da indexação e recuperação da informação 83
Gráfico 12 - Orientações sobre exaustividade nos repositórios 84
Gráfico 13 - Tradução dos termos para outros idiomas. 86
Gráfico 14 - Resistência dos autores em relação à tradução conceitual. 89
Gráfico 15 - Dificuldades enfrentadas na indexação de pesquisas referentes à Região Norte/Amazônia. 90
Gráfico 16 - Dificuldades enfrentadas na indexação de pesquisas referentes à Região Norte/Amazônia. 92

LISTA DE SIGLAS E ABREVIAÇÕES

AD – Análise Documentária

AL – Acesso Livre

BDTD - Biblioteca Digital Brasileira de Teses e Dissertações

BN – Biblioteca Nacional

BOAI - Budapeste Open Access Initiative

CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CDD – Classificação Decimal de Dewey

CDU – Classificação Decimal Universal

CETEC – Fundação Centro Tecnológico de Minas Gerais

CI – Ciência da Informação

CNPq – Conselho Nacional de Pesquisa

CTI Renato Archer – Centro de Tecnologia da Informação Renato Archer

DCMI - Dublin Core Metadata Initiative

DECS – Descritores em Ciências da Saúde

FINEP – Financiadora de Estudos e Projetos

FSA – Fundação Santo André

FUNCATE - Fundação de Ciências, Aplicações e Tecnologia Espaciais

FURG – Fundação Universidade Federal do Rio Grande

HP - Hewlett-Packard

IBICT – Instituto Brasileiro de Ciência Tecnologia e Informação

IEN – Instituto de Energia Nuclear

INMETRO - Instituto Nacional de Metrologia Normalização e Qualidade Industrial

INT – Instituto Nacional de Tecnologia

LDs – Linguagens Documentárias

LN – Linguagem Natural

MCT - Ministério da Ciência e Tecnologia

MeSH – Medical Subject Headings

MIT Libraries – Instituto de Tecnologia de Massachusetts

MPEG – Museu Paraense Emílio Goeldi

MTD3-BR - Padrão Brasileiro de Metadados para Descrição de Teses e Dissertações

MTD-BR - Padrão Brasileiro de Metadados para Descrição de Teses e Dissertações

NDLTD - Network Digital Library of Thesis and Dissertation (NDLTD)

NIH - U.S. National Library of Medicine

OAI-PMH - Open Archives Initiative Protocol for Metadata Harvesting

OC – Organização do Conhecimento

OI – Organização da Informação

PCAL - Portal de Publicação Seriada de Acesso Livre

PDU – Plano Diretor

PPGCI – Programa de Pós-Graduação em Ciência da Informação

PUC-PR – Pontifícia Universidade Católica do Paraná

PUC-RIO - Pontifícia Universidade Católica do Rio de Janeiro

RC – Representação do Conhecimento

RepositoriUM – Repositorio da Universidade do Minho

Ris - Repositórios Institucionais

RIUFPA – Repositório Institucional da UFPA

SOCS – Sistemas de Organização do Conhecimento

TE – Termo Específico

TG – Termo Geral

TICs – Tecnologias da Informação

TTI – Tratamento Temático da Informação

UEPG – Universidade Estadual de Ponta Grossa

UFAC – Universidade Federal do Acre

UFAL – Universidade Federal de Alagoas

UFAM – Universidade Federal do Amazonas

UFBA – Universidade Federal da Bahia

UFF - Universidade Federal Fluminense

UFG – Universidade Federal de Goiás

UFGD – Universidade Federal da Grande Dourados

UFMA – Universidade Federal do Maranhão

UFMT – Universidade Federal do Mato Grosso

UFOP – Universidade Federal de Ouro Preto

UFOPA – Universidade Federal do Oeste do Pará

UFPA – Universidade Federal do Pará

UFPE – Universidade Federal de Pernambuco

UFPEL – Universidade Federal de Pelotas

UFRA – Universidade Federal Rural da Amazônia

UFRGS – Universidade Federal do Rio Grande do Sul

UFRN – Universidade Federal do Rio Grande do Norte

UFRR – Universidade Federal de Roraima

UFS – Universidade Federal de Sergipe

UFSC - Universidade Federal de Santa Catarina

UFSCAR – Universidade Federal de São Carlos

UFT – Universidade Federal do Tocantins

UFU – Universidade Federal de Uberlândia

UFV – Universidade Federal de Viçosa

UFVJM – Universidade Federal dos Vales do Jequitinhonha e Mucuri

UNB – Universidade de Brasília

UNIFAP – Universidade Federal do Amapá

UNIFESSPA – Universidade Federal do Sul e Sudeste do Pará

UNIR – Universidade Federal de Rondônia

USCS - Universidade Municipal de São Caetano do Sul

USP - Universidade de São Paulo

UTFPR - Universidade Tecnológica Federal do Paraná

WWW – World Wide Web

SUMÁRIO

1 INTRODUÇÃO 17

2 ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO: LINGUAGENS DOCUMENTÁRIAS 23

2.1 Organização e representação do conhecimento 23

2.2 O processo de indexação e as linguagens documentárias alfabéticas 26

3 REPOSITÓRIOS INSTITUCIONAIS UNIVERSITÁRIOS: UM PANORAMA BRASILEIRO 37

3.1 O acesso livre e os repositórios digitais 37

3.1.1 Acesso aberto no Brasil e repositórios institucionais brasileiros 42

3.2 Estrutura dos repositórios institucionais universitários 46

3.2.1 O campo assunto nos repositórios institucionais universitários 48

4 METODOLOGIA 53

4.1 Descrição do universo da pesquisa e os sujeitos da pesquisa 53

4.2 Instrumento de coleta de dados: questionário 54

4.3 Elaboração de categorias de análise 55

4.4 Forma de análise dos dados 56

5 ANÁLISE DOS DADOS 58

5.1 Seção 1: identificação dos repositórios e das atividades dos indexadores 58

5.1.1 Identificação dos repositórios participantes da pesquisa 58

5.1.2 Análise das atividades dos indexadores 61

5.2 Seção 2: política de indexação e uso das linguagens documentárias 65

5.2.1 Adoção de política de indexação 65

5.2.2 Uso de linguagem documentária, linguagem natural ou ambas 67

5.2.3 Indexação de documentos digitais 71

5.3 Seção 3: Tradução conceitual dos termos de indexação para linguagens documentárias 74

5.3.1 Análise de assunto 74

5.3.2 Tradução conceitual 79

5.4 Discussão geral dos resultados 93

6 CONSIDERAÇÕES FINAIS 97

REFERÊNCIAS 99

APÊNDICE 107

APÊNDICE A – Questionário: Utilização das Linguagens Documentárias (LDs) na tradução conceitual em repositórios da Região Norte do Brasil 108

1 INTRODUÇÃO

Esta pesquisa teve como tema central o uso de Linguagens Documentárias (LDs) na representação de documentos disponibilizados em repositórios digitais, mais especificamente os repositórios institucionais mantidos pelas Universidades Federais da Região Norte do Brasil.

Representar um documento consiste em estabelecer termos que sintetizem o(s) assunto(s) contido(s) nele, de forma a orientar o pesquisador quanto à escolha de sua pertinência ou não pertinência, sem ter que ler o documento inteiro. A representação é uma etapa fundamental do processo de análise documentária, este visa extrair dos documentos os seus assuntos mais relevantes, indexando-os e tornando esses documentos encontráveis dentro de um sistema de informação. A indexação, segundo Cesarino e Pinto (1978), é uma atividade subjetiva, de cunho intelectual, ou seja, não pode ser mecânica e estática, depende da objetividade de cada catálogo ou biblioteca, além também da percepção e conhecimento do indexador. Dessa forma, embora em um primeiro momento a indexação pareça uma tarefa simples, ela não pode ser vista ou reduzida a tal ideia, como reforçam Cesarino e Pinto (1978) ao citarem que ela é, na verdade, uma atividade complexa, pois, caso um documento não seja apropriadamente representado, a informação ali contida pode se perder, impossibilitando, dessa maneira, sua recuperação.

Sobre a afirmativa acima, Chaumier (1988, p. 63) considera que uma indexação ineficaz é capaz de proporcionar "ruídos" (documentos que não servirão ao objetivo da pesquisa) e "silêncios", quando a pesquisa não retorna resultados; desta maneira, é fundamental que o indexador perceba a importância do seu papel ao analisar e representar um documento. Sendo então a análise documentária o processo pelo qual o indexador toma conhecimento dos assuntos tratados no texto, lendo-o de maneira técnica, chegando assim à sua síntese. Somente após a síntese, é possível escolher os termos que o representarão o documento para, enfim, traduzi-los para uma Linguagem Documentária (LD), que é uma linguagem que tem por objetivo padronizar os termos utilizados na indexação.

As características das LDs se modificaram ao longo do tempo, assim como a finalidade para a qual elas foram desenvolvidas, essa evolução as levou de simples listas de termos, com o objetivo de organizar documentos, a estruturas mais complexas, ramificadas e aprofundadas hierárquica e semanticamente. Isto se deu pela necessidade de se representar de maneira mais específica o conteúdo dos documentos, pois o foco das bibliotecas não era mais a sua simples organização para as suas múltiplas possibilidades de uso (NOVELLINO, 1996). Ou seja, o foco não era mais apenas uma organização simplória dos livros, mas sim a

recuperação de seus conteúdos. Pensando nesse aspecto de recuperabilidade, isto torna-se um desafio ainda maior quando falamos de recuperação de informação em ambientes web, pois o volume de documentos digitais é muito grande, exigindo, assim, uma indexação mais precisa, que proporcione uma rápida recuperação.

Lidar com o problema da recuperação de informações é uma das preocupações da Ciência da Informação (CI), que tem por premissa fundamental lidar com o fluxo informacional, desde sua gênese até seu uso; sendo que as tecnologias da informação potencializaram sua produção e seu acesso, diversificando os tipos de suporte, como já assinalava Bush (1945) ao propor que os suportes mecânicos teriam repercussão nos registros científicos. Bush (1945) também sugere a compressão através das tecnologias, ou seja, a possibilidade de registro de um grande volume de informação em um único e reduzido suporte. Essas "visões" elucubradas por ele puderam se tornar realidade com o advento da informática e dos registros digitais e, também, com o nascimento da internet. Dessa maneira, foi o desenvolvimento das tecnologias digitais que possibilitou ao homem criar espaços que abrigam enormes volumes de informação, disponibilizando-os para acesso a partir de qualquer parte do planeta.

Foram essas tecnologias que possibilitaram o nascimento dos repositórios digitais, que são ferramentas tecnológicas originadas no seio do movimento Acesso Livre (AL). O AL é uma iniciativa que tem por objetivo democratizar o acesso às publicações científicas, resgatando-as do controle dos periódicos tradicionais (KURAMOTO, 2012). Uma das razões para o desenvolvimento do AL foi o monopólio exercido pelas grandes editoras sobre a divulgação das pesquisas e o alto preço cobrado por elas pelos exemplares, o que dificultava o acesso, tanto pelo fator financeiro quanto pelo fator tempo, para que o pesquisador tivesse o documento em mãos. Assim, fez-se necessário a criação de instrumentos que possibilitassem uma democratização desse acesso, tanto para pesquisadores, quanto para a comunidade em geral, pondo à disposição de todos os resultados das pesquisas, principalmente aquelas financiadas com recursos púbicos (ROSA; GOMES, 2010, p. 153, KURAMOTO, 2006, p. 92).

Colocar à disposição da sociedade os documentos originados de pesquisas financiadas com dinheiro público torna-se uma questão de justiça, visto que se o dinheiro vem dos cofres públicos, nada mais justo que a sociedade tenha acesso ao que está financiando. Porém, essa evolução na comunicação científica só foi possível através do armazenamento eletrônico e em nuvem, o que proporcionou, através da web, acesso amplo aos documentos.

Essa amplitude de acesso, no entanto, só se concretiza através da recuperação relevante de conteúdo que, por sua vez, está intimamente ligada à representação temática dos documentos no momento em que dão entrada nos sistemas informacionais. Assim, os termos (palavras-chave) não só devem representar os assuntos contidos no texto, mas também têm a missão de torná-lo visível, encontrável no meio da enxurrada de informação existente no mundo virtual. Nesse ambiente, um documento bem representado será mais facilmente encontrado e utilizado, pois o usuário nem sempre tem tempo para analisar seu conteúdo de forma mais aprofundada, assim, um documento que poderia lhe ter sido útil acaba se tornando "invisível".

Dar visibilidade aos documentos, aos autores e às instituições é um dos principais objetivos dos repositórios, pois, ao disponibilizar suas pesquisas na internet através dos repositórios, as universidades reforçam ainda mais seu papel de produtoras de conhecimento. Nesse contexto, é importante a utilização de LDs nos repositórios institucionais universitários, já que elas podem impedir a dispersão documentária, isto é, diversos documentos que tratam do mesmo tema, sendo indexados sob termos diferentes. A qualidade da indexação mostra o comprometimento da instituição com os pesquisadores, pois demonstra que o documento não foi colocado ali de qualquer forma e que há uma preocupação com a disponibilização não apenas do objeto em si, mas do conteúdo que ele trata.

Disponibilizar conteúdo de forma *online* e gratuita, em diversos formatos (*jpg*, *pdf*, *word*, etc.) é um dos fatores que tornou os repositórios tão populares entre as universidades, que precisam fazer com que a roda do conhecimento gire de forma rápida e eficaz. Os repositórios proporcionaram isso: eficácia na disseminação da informação. Além de também impulsionarem a visibilidade da instituição e dos pesquisadores, pois, sendo interoperáveis, eles possibilitam a captação de informações de qualquer parte do mundo, formando uma rede.

Bush (1945) já percebia que a amplitude de acesso aos documentos se daria através do uso de tecnologias, como máquinas e instrumentos capazes de armazenar um grande volume de informação, tornando-o disponível para posterior recuperação. Mooers (1951, p. 25) define a recuperação da informação como um "processo ou método pelo qual um usuário em potencial de informação é capaz de converter sua necessidade de informações em uma lista real de citações para documentos em armazenamento contendo informações útil para ele" (Tradução nossa). Ou seja, recuperar a informação é trazer à tona documentos de uso relevante para o pesquisador, que o seleciona dentre tantos outros disponíveis.

Diante do exposto, a utilização das LDs por parte de bibliotecas e centros de documentação é uma preocupação constante, como asseguram Fujita et al. (2019), discutindo o uso dessas linguagens no contexto de bibliotecas universitárias. Outros autores como Chaumier (1988), Pinto (1985), Cesarino e Pinto (1978), também discutem o processo de indexação e a utilização das LDs, já que elas adquiriram um papel importante na tradução dos termos selecionados dentro dos textos. Sobre a recuperabilidade da informação na web, tem-se trabalhos como os de Miranda (2005, 2006), onde o autor discute a organização e a representação do conhecimento em ambientes virtuais, analisando a relação da linguagem natural e os motores de busca. Há também trabalhos direcionados para a web semântica, como artigo publicado por Souza e Alvarenga (2004), em que é discutida a relação da Ciência da Informação (CI) e a representação de documentos digitais. Assim, mesmo diante dos estudos voltados para a indexação em ambientes digitais e virtuais, ainda são poucos os trabalhos que se voltam para a o uso de LDs em repositórios digitais, instrumentos com crescente adoção por parte das instituições de pesquisa como forma de organizar e disseminar a informação.

Dessa forma, cabe investigar como a organização e a representação do conhecimento, assim como o uso das LDs, contribuem para a organização da informação dentro do cenário dos repositórios, mais especificamente, os repositórios institucionais universitários, que comportam amplo volume de pesquisa. Tomando como objeto de estudo o uso das LDs em repositórios digitais, indaga-se: como as LDs vêm sendo utilizadas pelos profissionais indexadores em repositórios institucionais universitários federais na região norte do Brasil?

A **proposta** é traçar um panorama sobre a utilização de linguagens documentárias na indexação de documentos nos repositórios digitais das Universidades Federais da Região Norte do Brasil a partir da aplicação de questionários com os indexadores. Assim, diante do exposto, o **objetivo geral** desta pesquisa é contribuir para estudos sobre a organização e a representação do conhecimento em repositórios institucionais universitários, a partir de investigação sobre o uso de linguagens documentárias em repositórios de Universidades Federais da Região Norte do Brasil. Dessa forma, para nortear os passos desta pesquisa e atingir o objetivo maior exposto acima, foram traçados alguns objetivos específicos, os quais estão descritos abaixo:

a) buscar na teoria conceitos sobre o processo de indexação na organização e representação do conhecimento;

b) buscar e compilar reflexões teóricas que explorem elementos históricos e características gerais dos repositórios digitais, assim como a evolução dos repositórios digitais universitários no Brasil;
c) investigar o uso das linguagens documentárias alfabéticas no processo de indexação em repositórios a partir da aplicação de questionários nos repositórios das Universidades Federais da Região Norte do Brasil.

Utilizou-se, como metodologia, de uma abordagem qualitativa para a investigação do uso das linguagens especializada para indexação de documentos em repositórios digitais, configurando-se também como um estudo de caso, ao ter como universo de pesquisa os repositórios universitários federais da Região Norte do Brasil. Os sujeitos da pesquisa foram 8 bibliotecários (um por repositório), aos quais foi enviado um link de acesso ao instrumento de coleta de dados, caracterizado por um questionário semiaberto online, dividido em três macros sessões de análise e 23 questões direcionadas aos aspectos considerados importante a serem observados.

O trabalho está dividido em quatro capítulos, as quais estão descritas abaixo:

Capítulo 1: dedicado à introdução do tema, faz um panorama sobre a indexação e uso das LDs, assim como o contexto de nascimento dos repositórios digitais e a importância da padronização de termos em ambiente web.

Capítulo 2: explana sobre a Organização do Conhecimento (OC) e sua contribuição para a elaboração das LDs, que hoje são linguagens baseadas na terminologia de uma área, proporcionando a padronização do uso de termos indexadores.

Capítulo 3: é abordado as LDs de forma mais específica, tais como suas características, funções e estruturas, bem como a importância que a política de indexação tem diante do estabelecimento de um padrão de uso das LDs nos repositórios institucionais universitários.

Capítulo 4: é dedicado à metodologia, descrevendo o universo da pesquisa, a amostra, assim como os instrumentos utilizados para a coleta de dados.

Capítulo 5: dedicado à análise dos dados.

Capítulo 6: dedicado às considerações finais diante dos fatos observados durante a pesquisa.

Estando o trabalho distribuído dessa forma, o quadro 1 apresenta a relação entre os capítulos e os objetivos propostos:

Quadro 1 - Relação dos objetivos com os capítulos da pesquisa

Objetivos específicos	Capítulos relacionados
a) buscar na teoria conceitos sobre o processo de indexação na organização e representação do conhecimento;	ılo 2: Organização e representação do conhecimento: linguagens documentárias como instrumentos de indexação.
b) buscar e compilar reflexões teóricas que explorem elementos históricos e características gerais dos repositórios digitais, assim como a evolução dos repositórios digitais universitários no Brasil;	Capítulo 3: Repositórios institucionais universitários: um panorama brasileiro.
c) investigar o uso das linguagens documentárias alfabéticas no processo de indexação em repositórios a partir da aplicação de questionários nos repositórios das Universidades Federais da Região Norte do Brasil.	Capítulo 4: Metodologia Capítulo 5: Análise dos dados Capítulo 6: Considerações finais

Fonte: elaborado pela autora, 2019.

2 ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO: linguagens documentárias como instrumentos de indexação

A Organização do Conhecimento (OC) é a base para a construção de Linguagens Documentárias (LDs), estas, por sua vez, são instrumentos de padronização na atividade de representação da informação. Dessa forma, para esta pesquisa, optou-se em trabalhar com a conceituação teórica sobre OC e os Sistemas de Organização do Conhecimento (SOCS), assim como também das principais LDs utilizadas na indexação de documentos.

2.1 Organização e representação do conhecimento

Ao longo de sua história, o homem buscou relacionar-se e entender o mundo de forma a garantir sua sobrevivência, desenvolvendo técnicas e instrumentos que facilitassem sua vida, assim como dedicou-se a compreender os fenômenos que aconteciam ao seu redor. Enquanto predominava o teocentrismo (Deus é o centro de tudo), a relação do homem com o mundo, assim como com sua consciência, era explicada segundo uma visão divina, ou seja, o homem entendia o mundo de acordo com explicações divinas, mitológicas e sobrenaturais. A Filosofia nasceu como um contraponto dessa visão, pois buscava compreender a relação do homem com ele mesmo e com o mundo de forma observada, lógica, racional e sistemática, por meio de experimentações, deixando de lado as explicações meramente mitológicas para obter conhecimento.

Conhecimento, no verbete do dicionário, é definido como “ato ou efeito de conhecer”, “conjunto de informações e princípios que o homem aprendeu”, “o ato de conhecer por meio da razão e/ou experiência” (MICHAELIS..., 2015), muito embora haja outras definições, no sentido deste estudo, tornam-se mais interessantes essas três. Assim, de uma maneira simples, podemos notar que a definição de conhecimento é a consciência que o ser humano tem do mundo em que vive, aquilo que ele conhece seja através dos sentidos ou da razão.

Ver e explicar o mundo através da razão, lhe dando um sentido, não é uma tarefa simples, assim, se torna necessário o estabelecimento de métodos e tais métodos têm suas bases na Filosofia, quando as principais correntes filosóficas dizem que o conhecimento pode ser adquirido utilizando-se os sentidos ou a razão. Os filósofos que acreditavam que a única maneira de se conhecer o mundo era através dos sentidos e da experiência eram chamados de empiristas, a exemplo de Bacon, Locke e Hobbes; já os que acreditavam que o

conhecimento era fruto unicamente da razão eram chamados de racionalistas, a exemplo de Pascal, Spinoza e Descartes.

Posteriormente a estas duas linhas teóricas sobre conhecimento, já dentro da considerada filosofia moderna ou contemporânea (JAPIASSU; MARCONDES, 2001; SILVEIRA, 2002), Kant formula uma outra teoria, a filosofia transcendental (SILVEIRA, 2002), na qual ele acredita que nem o racionalismo, nem o empirismo são completamente confiáveis para levar o homem ao conhecimento. O que Kant acredita é que a razão e os sentidos são complementares, portanto, para adquirir conhecimento, o homem não pode usar exclusivamente da razão ou confiar apenas nos sentidos e experiências, devendo utilizar-se dos dois.

A Filosofia nos mostra que conhecimento existe e se desenvolve de forma abstrata, dentro da mente humana, o conhecimento é tudo o que percebemos e absorvemos do mundo e, embora não seja um objeto palpável, ele pode, no entanto, ser fixado, transformando-se em informação. Le Coadic (2004, p. 5) corrobora com essa afirmação ao dizer que a informação pode ser definida como "um conhecimento inscrito (gravado) sob a forma escrita (impressa ou numérica), oral ou audiovisual". Dessa forma, podemos entender que o conhecimento gera a informação e informação, quando absorvida, é transformada em novo conhecimento. Todavia, esse ciclo depende da percepção do receptor, o qual interpretará a mensagem e associará a ela um valor (BRASCHER; CAFÉ, 2008), dando prosseguimento ao ciclo, gerando novos conhecimentos e informações.

Assim, tomando como princípio que o conhecimento é algo abstrato, uma questão vem à nossa mente: se o conhecimento não é um objeto palpável, como então organizá-lo? Essa questão encontra resposta em Dahlberg (1978), onde a autora cita o uso da linguagem como forma de representação e comunicação dos conhecimentos gerados pela humanidade. A linguagem oferece os signos que, por sua vez, são dotados de um significante (imagem acústica) e um significado (conceito) (SAUSSURE, 2006, p. 79) e foi sobre este segundo elemento que Dahlberg (1978) elaborou sua teoria. Ela nos traz a definição de conceito como uma "compilação de enunciados verdadeiros sobre determinado objeto, fixada por um símbolo linguístico" e, dessa forma, representando os conceitos através de signos da linguagem, é possível estabelecer relações entre eles. Assim, a essas relações abstratas, que dividem o conhecimento em áreas específicas, levando em consideração suas características, hierarquia e funções, chamamos Organização do Conhecimento (OC). Ao se trazer o conhecimento para o mundo físico (dos sentidos), eles passam a ser representados por

elementos do mundo real, como os signos linguísticos, a este processo, chamamos Representação do Conhecimento (RC). A RC, segundo Brascher e Café (2008, p. 8), se faz

> por meio de diferentes tipos de sistemas de organização do conhecimento (SOC) que são sistemas conceituais que representam determinado domínio por meio da sistematização dos conceitos e das relações semânticas que se estabelecem entre eles.

Isso nos demonstra que o conhecimento (abstrato) pode ser traduzido para o mundo real (dos sentidos) e, de forma estruturada, oferecer instrumentos que, por sua vez, permitem que o conhecimento registrado (informação) seja organizado.

Sobre os SOCS, Carlan (2010, p. 28) afirma que eles se configuram como

> [...] instrumentos que fazem a tradução dos conteúdos dos documentos originais e completos, para um esquema estruturado sistematicamente, que representa esse conteúdo, com a finalidade principal de organizar a informação e o conhecimento e, consequentemente, facilitar a recuperação das informações contidas nos documentos.

A partir desta definição, pode-se perceber que os SOCS são sistemas lógicos, construídos com o objetivo de organizar o emaranhado de conceitos e relações conceituais dentro de uma área, a fim de que a informação por ela produzida possa ser organizada e recuperada.

A ideia dos SOCS como estruturas destinadas a organizar a informação tendo por finalidade a sua recuperação também está presente em Hodge's (apud HJORLAND, 2015) e Carlan e Medeiros (2011), podendo, assim, mostrar que a preocupação com a recuperação da informação deve se fazer presente já quando o documento dá entrada no sistema, no momento de sua representação através de um SOC.

Para Carlan (2010), o termo SOC é uma denominação mais moderna para os instrumentos que auxiliam na representação da informação, visto que eles abrangem não somente as tradicionais LDs, como os cabeçalhos de assunto e os tesauros, mas também as linguagens mais modernas, utilizadas para a organização da informação em sistemas digitais. Como o objetivo principal deste trabalho é observar a indexação em repositórios digitais mantidos por bibliotecas tradicionais, a classe de SOC utilizada será a das LDs no processo de tradução conceitual e representação da informação.

O processo de representação da informação pode se dar de maneira temática ou descritiva, sendo que a representação temática visa representar, através de termos padronizados pelas LDs, o tema tratado; enquanto que a representação descritiva visa descrever a estrutura física do documento (MAIMONE; SILVEIRA; TÁLAMO, 2011, p.

28). Para as autoras, a representação da informação tem como principal função mediar o relacionamento documento/usuário, assim, representar um documento significa atribuir-lhe termos que traduzam seu conteúdo de forma sintetizada e, para que a relação documento/usuário seja eficiente, é necessário a padronização destes termos.

Dito isso, destaca-se que a OC, ao reunir termos sob certo domínio e possibilitar a estruturação das linguagens documentárias, oferece importantes instrumentos à atividade de indexação. Dessa forma, é preciso compreender as fases desse processo e também de que maneira se dá o uso destas linguagens, questões que veremos no tópico a seguir.

2.2 O processo de indexação e as linguagens documentárias alfabéticas

Falar de indexação requer um cuidado especial, tanto para visualizar seus caminhos, quanto para compreender sua conceituação, pois existem, dentro da CI, divergências quanto à visão que autores e pesquisadores adotam em relação a ela. Há aqueles que a tomam como um processo (corrente inglesa) e aqueles que a tem como uma etapa dentro de um outro processo maior (corrente francesa). Dessa maneira, para chegarmos ao uso das LDs, precisamos entender o conceito de indexação e os processos que a compõem, como a fase onde se analisa o texto a fim de extrair conceitos e a fase da tradução conceitual (a qual é o ponto principal deste trabalho), onde os termos serão traduzidos e padronizados por uma LD.

O termo indexação nos remete ao índice que, por sua vez, é a indicação dos assuntos mais importantes tratados em um livro, em ordem alfabética relacionando-os à página onde aparecem, facilitando, assim, sua localização (MICHAELIS...). Para a CI, indexação é o ato de representar os assuntos de um documento através de termos, visando sua recuperação (CESARINO; PINTO, 1978). O tema da indexação, dentro das definições teóricas desenvolvidas pela CI, é entendido de diferentes maneiras, sendo as principais correntes desenvolvidas por teóricos franceses e ingleses, como já mencionado anteriormente. A corrente francesa entende a indexação como uma etapa dentro de um processo maior, denominado Análise Documentária (AD); já a corrente inglesa entende a indexação como o próprio processo (SILVA; FUJITA, 2004; GUIMARÃES, 2009).

Essa diferença de abordagens se dá, segundo Guimarães (2009), devido ao desenvolvimento metodológico e instrumental que a atividade de Tratamento Temático da Informação (TTI) veio sofrendo ao longo dos anos, evolução impulsionada, principalmente, pelo aumento do volume de informações e a necessidade de se recuperar além dos aspectos óbvios do documento. Inicialmente, a atividade de indexar um documento se centrava na

elaboração de índices e catálogos, objetivando a organização dos documentos/livros. Quando a percepção do TTI se voltou para o contexto e o uso do documento, preocupando-se com sua recuperação, outros processos foram sendo incorporados, como a síntese e conceituação do(s) assunto(s) através de termos (GUIMARÃES, 2009).

A indexação, portanto, segundo a corrente francesa, se configura como um grupo de atividades que se relacionam entre si, visando extrair o conteúdo dos documentos através de uma análise conceitual, para, em seguida, representá-los através de termos, objetivando sua posterior recuperação (GARDIN, 1974 apud BENTES PINTO; PINTO, 1985). Desta maneira, este trabalho se alinha com a corrente francesa em sua fundamentação teórica, compreendendo a indexação como uma parte da AD, esta, por sua vez, segundo Cunha (1987, p. 38), constitui-se por "um conjunto de procedimentos efetuados com o fim de expressar o conteúdo de documentos, sob formas destinadas a facilitar a recuperação da informação". Nota-se aqui, mais uma vez, que a recuperação é objetivo final do tratamento da informação, sendo assim, Guimarães (2009) categoriza a AD em duas atividades fundamentais, a elaboração de resumos e a indexação.

Para Chaumier (1988, p. 63), a indexação "é a parte mais importante da análise documentária" e, baseado em Van Slype (1977), ele expõe que ela é composta por quatro ações, sendo elas: conhecer o conteúdo dos documentos; escolher os termos para representação; tradução dos termos escolhidos para uma linguagem documentária; e, incorporar os elementos sintáticos. Fujita (2003), simplificando as ações dentro da indexação, propõe que ela é composta por dois estágios: o primeiro é o analítico, que acontece quando se executa a compreensão do texto, identificando e selecionando conceitos que mais tarde passarão pelo segundo estágio, que é a tradução para uma linguagem documentária. Pode-se notar que, para ambos os autores, a indexação depende do conhecimento que o responsável por ela busca ter do texto, escolhendo adequadamente os termos que representem seu conteúdo para posterior recuperação. Assim, baseado em estudos como o de Chaumier (1988) e Redigolo (2010; 2014), cabe expressar e entender cada uma das etapas que compõem o processo de análise documentária, sendo elas: leitura documentária, identificação de conceitos e seleção de conceitos. A seguir, uma breve conceituação de cada uma destas etapas:

- A leitura é a primeira fase da análise documentária e visa trazer ao conhecimento do indexador o(s) conteúdo(s) do(s) documento(s). A leitura documentária caracteriza-se como uma atividade técnica, realizada com fins

específicos, que é extrair conceitos que representem o conhecimento registrado. Para Fujita (2004), este tipo de leitura tem caráter profissional, pois faz parte das atividades atribuídas ao indexador, objetivando a sintetização do conteúdo do documento para então representá-lo. Fujita (2004) identifica três variáveis que influenciam o processo de leitura com fins documentários, são elas: o texto, o leitor e o contexto. O texto possui uma estrutura lógica, onde podem ser buscadas pistas para identificação dos assuntos; já o leitor, além de conhecer essa estrutura, deve também ter conhecimento prévio da área onde ele (texto) se insere. O contexto refere-se a elementos como as condições físicas do espaço/local onde se desenvolve a atividade.

- A identificação de conceitos é a etapa que se segue após a leitura documentária, é neste momento que serão identificados os assuntos dos quais trata o documento. Sobre isso, Dias e Naves (2013, p. 58) afirmam que a identificação de conceitos "tem como produto um assunto, que representa o conteúdo informacional de um texto". Com o intuito de direcionar a identificação de conceitos, a Associação Brasileira de Normas Técnicas (1992, p. 2), através da NBR 12676/1992, especificou algumas questões a serem observadas, são elas:

    - qual o assunto de que trata o documento?
    - como se define o assunto em termos de teorias, hipóteses, etc.?
    - o assunto contém uma ação, uma operação, um processo?
    - o documento trata do agente dessa ação, operação, processo, etc.?
    - o documento se refere a métodos, técnicas e instrumentos especiais?
    - esses aspectos foram considerados no contexto de um local ou ambiente especial?
    - foram identificadas variáveis dependentes ou independentes?
    - o assunto foi considerado sob um ponto de vista interdisciplinar? (p. ex. um estudo sociológico da religião)

- Essas indagações podem servir de norte ao indexador no que se refere à identificação de conceitos pertinentes dentro do documento, estabelecendo, assim, uma lista de termos que poderão ser utilizados na representação deste. É a seleção de conceitos que determinará qual o termo que representa o assunto principal, assim como os assuntos secundários e também os termos

qualificadores, que proporcionarão uma maior especificidade e exaustividade na indexação. É nesta etapa que se evidencia um elemento já citado anteriormente por Fujita (2004), quando a autora diz que faz-se necessário que o indexador tenha conhecimento do contexto onde o documento está inserido. O termo "informação", por exemplo, é entendido de diferentes formas por bibliotecários, jornalistas, biólogos, etc. Para o bibliotecário, a informação é um conjunto de dados dotados de sentido, seja ela uma imagem, um texto, um som; para os jornalistas, informação é a matéria veiculada pelos meios de comunicação; e para os biólogos, informação é uma mensagem trocada entre elementos biológicos como células, moléculas, órgãos. Por estas razões é que o indexador deve ter bom conhecimento da área onde o documento está inserido, pois este pode servir de diferentes formas a públicos diferentes. É através deste elemento que o indexador decide quais conceitos são importantes representar para que o documento seja utilizado na área em que a biblioteca está inserida.

- **seleção de conceitos:** A seleção é a terceira etapa da análise documentária e consiste na escolha, pelo indexador, dos termos preferidos para representar os documentos. Sobre tal processo, Redigolo (2010, p. 75) afirma que

> Uma vez identificados os conceitos que representam a informação de um documento, é necessário que o catalogador selecione termos de descrição, palavras ou frases que sirvam para representar exatamente o conceito identificado e que sejam mais adequados para a recuperação.

Isto quer dizer que na escolha dos termos para a representação do conteúdo do documento, é necessário considerar seu consequente uso, ou seja, escolher os termos que tornarão o documento recuperável dentro do sistema em que está inserido. A seleção de conceitos demanda bom senso do indexador, para que os termos escolhidos a comporem o índice do catálogo sejam os que melhor refletem a informação que interessa ao público do sistema de informação.

2.2.1 Tradução conceitual e linguagens documentárias

Após a leitura, a identificação e a seleção de conceitos, a próxima etapa da análise documentária é tradução conceitual, onde os termos selecionados dentre aqueles

identificados no documento passarão da LN para uma LD. A tradução conceitual, segundo Lancaster (2004), é o processo onde os termos selecionados serão transpostos para um "um conjunto de termos de indexação", ou seja, os termos escolhidos serão traduzidos para termos padronizados por uma LD e, dessa forma, indexados para posterior recuperação.

A recuperação é um termômetro da indexação, pois ela permite medir se as estratégias e os instrumentos adotados estão sendo satisfatórios, ou seja, se a biblioteca está conseguindo entregar a informação mais precisa ao usuário. O aparecimento de ruídos (documentos não necessários ao interesse do pesquisador) e silêncios (a pesquisa não retorna documentos úteis) está intrinsicamente ligado à qualidade da indexação e a qualidade da indexação está ligada às etapas anteriores do processo: leitura, seleção de conceitos e tradução em uma LD (CHAUMIER, 1988). Todas essas etapas devem estar claras para o indexador, assim como ele também deve entender e estar comprometido com o sistema para o qual trabalha, pois as diretrizes e os instrumentos terão pouca serventia se não forem utilizados de maneira adequada. Assim, a adoção de LDs por parte das bibliotecas é de suma importância, já que se os documentos disponíveis ali tiverem pouco ou nenhum uso, isso pode ser uma consequência da indexação inadequada.

Para entender melhor o que são as LDs, tomemos como base as ideias presentes em Cintra et al. (2002) que definem as LDs como linguagens construídas com base na linguagem natural (LN) e na linguagem especializada, buscando a padronização na representação de assuntos de um documento, visando sua posterior recuperação. Nota-se aqui, mais uma vez, que a recuperação deve ser o foco principal de um sistema de informação e, nesse cenário, a padronização dos termos utilizados na representação assume um papel crucial. Porém, não basta apenas compreender que as LDs são instrumentos de OC e representação da informação, é necessário entendê-las mais profundamente, buscando informações a respeito de seu funcionamento e estrutura.

A estrutura de uma LD, como já dito anteriormente, é apoiada em esquemas linguísticos de relacionamento, como na linguagem natural (LN). Contudo, no primeiro tipo de linguagem citado, o cuidado com a escolha dos termos e conceitos é maior, pois a ambiguidade deve ser eliminada e os sinônimos controlados dentro do domínio de uma área. Ainda, as relações existentes entre os conceitos estão presentes nos estudos de Dahlberg (1978) e, segundo a ela, podem se dar da seguinte forma:

- **a)** Lógica (baseada nas características comuns);
- **b)** Hierárquica (relação de pertinência entre os conceitos, do macro para o micro ou vice versa);

c) Partitiva (relação todo – parte);

d) Oposição (características opostas);

e) Funcionais (baseadas nos processos).

O relacionamento entre conceitos permite coordenar os termos e distribuí-los de forma lógica, possibilitando assim que a Documentação crie suas LDs de acordo com sua necessidade e uso. Uma LD, segundo Gardin (1968 apud CINTRA et al., 2002, p. 35), tem como características estruturais básicas: um léxico, uma rede paradigmática e uma rede sintagmática, a saber:

- **o léxico** pode ser entendido como o conjunto de termos relacionados a uma língua o que, no caso das LDs, diz respeito ao conjunto de termos relacionados a uma área do conhecimento;
- **a rede paradigmática** é a forma como os termos se relacionam e se associam por determinado contexto ou característica;
- **a rede sintagmática** é responsável pela relação lógica entre os termos.

Diante dessas características, podemos estabelecer uma LD como uma estruturação de termos que se relacionam entre si por suas características e conceitos e que podem ser colocados numa ordem lógica, que lhe dá sentido. Essa relação entre termos permite que, dependendo de seu uso nos sistemas de recuperação da informação, as LDs possam ser classificadas em pré e pós-coordenadas (CAVALCANTI, 1982), devido ao uso de qualificadores, visando uma maior especificidade, desse modo:

- **linguagens pré-coordenadas:** são construídas a partir de termos coordenados no momento da indexação, ou seja, quando dá a entrada no sistema, a informação já é colocada sob expressões (junção de termos) pré-determinados pelo indexador. Esse tipo de esquema deixa a busca um tanto engessada, não permitindo que o usuário faça suas próprias combinações, por isso são linguagens fechadas;
- **linguagens pós-coordenadas:** são linguagens que permitem que o usuário combine os termos que serão utilizados na hora da busca, ou seja, no momento da recuperação. Para Pinto (1985), a possibilidade de delegar ao usuário o relacionamento dos termos durante a busca é uma característica das LDs modernas, caracterizadas pelos sistemas automatizados de recuperação da informação.

Pré ou pós-coordenadas, modernas ou mais antigas, o fato é que as LDs desempenham um papel fundamental dentro do TTI, que é a padronização da representação

dos assuntos expressos nos documentos. As linguagens documentárias podem ser estruturadas de maneira sistemática, ou seja, de acordo com um sistema de grupos de termos, ligados semanticamente, ou de maneira alfabética. As linguagens documentárias alfabéticas podem ser classificam-se em:

a) **linguagens documentárias hierárquicas:** a exemplo da CDD (Classificação Decimal de Dewey) e da CDU (Classificação Decimal Universal), que são linguagens estruturadas de forma a abranger e representar todas as áreas do conhecimento. Uma característica das linguagens documentárias enciclopédicas é a generalidade de seus termos;
b) **linguagens documentárias alfabéticas:** constituídas a partir de termos de especialidade de determinada área do conhecimento ou de maneira geral, agrupando termos de todas as disciplinas. São estruturas hierarquizadas semanticamente, permitindo especificidades e aprofundamento na indexação.

O uso das linguagens hierárquicas está relacionado com o tema central do documento, determinando um conjunto de elementos que constituirão o número de chamada, o qual definirá a posição e a localização do documento no acervo. É o caso das já citadas CDD e CDU e também da tabela de Cutter, utilizada para distinguir uma obra entre as demais, utilizando uma combinação de letra e números baseada no sobrenome do autor e no título. Já as LDs alfabéticas, constituem estruturas organizadas com a finalidade de sintetizar e representar os assuntos relevantes na obra, através dos quais ela possa ser recuperada. Assim, tendo em vista o objetivo deste trabalho, que busca observar a indexação de documentos em repositórios institucionais universitários, serão utilizadas somente as LDs alfabéticas.

As primeiras linguagens documentárias alfabéticas a serem utilizadas foram as listas de cabeçalhos de assuntos, onde os termos preferidos para representar determinado conceito eram organizados, a fim de padronizar a entrada de documentos no sistema. Porém, devido a algumas limitações apresentadas pelos cabeçalhos de assunto, acabou se optando, posteriormente, pelo uso do vocabulário livre. No entanto, o uso de vocabulário livre também tinha suas lacunas e não conseguia satisfazer a necessidade de melhoria na recuperação da informação, pois o uso de vocabulário livre era campo aberto para a polissemia e a sinonímia entre os termos. Os cabeçalhos eram deficientes por não permitirem o desdobramento mais específico do conteúdo e os vocabulários livres não permitiam o estabelecimento de um padrão ou assunto principal (NOVELLINO, 1996). Assim, a busca por maior especificidade e padronização na indexação dos documentos levou à criação de instrumentos mais adequados de controle de vocabulário, os tesauros.

Tesauros e cabeçalhos de assunto são, ainda hoje, as LDs mais utilizadas pelas bibliotecas por atenderem ao objetivo destas, que é servir de mediadoras da informação, proporcionando sua recuperação e uso. Os cabeçalhos de assunto são definidos por Cesarino e Pinto (1978, p. 273) como "palavra ou grupo de palavras que expressam o conteúdo de um documento", ainda essas são, talvez, a "forma mais antiga de sistema alfabético".

Os cabeçalhos de assunto são um tipo de LD alfabética e pré-coordenada, assim, as primeiras linguagens desse tipo possibilitavam uma combinação limitada entre os termos, tanto pela extensão que o catálogo teria, quanto pelo trabalho que isso demandaria (CESARINO; PINTO, 1978). Ainda segundo essas autoras, a tomada de decisão sobre os termos autorizados era feitas pelos indexadores, tendo pouca ou nenhuma participação dos autores. Em um cabeçalho de assunto, a relação entre os termos é feita de forma hierárquica e alfabética, o que pode, de certa forma, prejudicar a aproximação de termos com conceitos semelhantes.

Dentre este tipo de LD, tem-se o *Library of Congress Subject Headings* (LCSH), lista de cabeçalhos de assunto criada e mantida pela Biblioteca do Congresso Americano. Durante anos o LCSH foi publicado de forma impressa, o que demandava bastante tempo para a pesquisa de termos. Hoje, o LCSH pode ser acessado exclusivamente online e possui aproximadamente 42 milhões de registros de autoridades entre as classes de sujeito, gênero e forma, abrangendo todas as áreas do conhecimento (LCSH). Embora tenha sido criado como um cabeçalho de assunto, o LCSH hoje já demonstra uma tesaurificação, como cita Fujita (2016) ao afirmar que a Library of Congress Subject Headings veio sofrendo alterações ao longo de sua história e hoje já adota estruturas mais complexas como a concordância semântica, o relacionamento hierárquico e lógico próprios dos tesauros.

Isso nos mostra que as LDs não são estáticas, elas evoluem e se adequam à necessidade informacional do usuário, observada pelos bibliotecários que, por sua vez, buscam a criação de novos instrumentos, assim como a atualização dos já existentes. Isso se dá através da observação da LN utilizada pelos autores que, em seu ciclo de produção de conhecimento, acabam por estabelecer, consensualmente, os termos que definem conceitos dentro de suas áreas, dando aos bibliotecários o aporte necessário para a construção e a atualização das LDs.

Foi neste contexto de necessidade informacional dos usuários e, consequentemente, necessidade dos bibliotecários, de terem em mãos instrumentos que melhor representem a informação, que nasceram os tesauros. Os tesauros são um tipo de LD que se encaixam na categoria de linguagens alfabéticas pós-coordenadas, pois permitem que o pesquisador reúna

os termos de acordo com sua pesquisa. Assim, são linguagens mais abertas, oferecendo maiores combinações entre os termos na hora da busca, ou seja, maior liberdade no uso de qualificadores.

Segundo a ANSI/NISO Z39.19 – 2005 (NISO, 2005, p. 30), um tesauro é "[...] um vocabulário controlado organizado em uma área do conhecimento e estruturado de forma que as diversas relações entre os termos sejam claramente mostradas e identificadas por indicadores padronizados de relacionamento". Os indicadores são expressões que conectam os termos, demonstrando suas relações, tais como: termo geral (TG), termo específico (TE), ver ou ver também.

São as relações semânticas que determinam a posição de um termo na teia conceitual, bem como quantas relações ele é capaz de oferecer dentro de uma terminologia, sendo que o controle terminológico é uma das principais características de um tesauro. Dessa forma, após expostas as características fundamentais das principais LDs utilizadas pelos indexadores, podemos traçar um paralelo sobre suas semelhanças e diferenças, o que pode ser observado a seguir:

Quadro 2 - Características de relacionamento entre os termos nas linguagens documentárias

LINGUAGEM DOCUMENTÁRIA	RELAÇÃO HIERÁRQUICA (do geral para o específico)	RELAÇÃO DE EQUIVALÊNCIA (use)	RELAÇÃO ASSOCIATIVA (ver/ver também)	RELAÇÃO SEMÂNTICA
TESAURO	Sim	Sim	Sim	Sim
LISTA DE CABEÇALHOS DE ASSUNTO	Sim	Sim	Não	Não
ONTOLOGIAS	Sim	Não	Não	Não
TAXONOMIAS	Sim	Não	Não	Não

Fonte: elaborado pela autora, 2019.

Como podemos notar no quadro 2, os tesauros são construções linguísticas mais complexas que os cabeçalhos de assunto, pois vão além da estruturação do macro ao micro ou de uma relação de "contém" e "está contido". Os tesauros ocupam-se também da relação entre conceitos e sentidos, formando uma verdadeira "árvore" de termos e são construções de abrangência mais estreita, dirigindo-se à terminologia de uma área específica do conhecimento, como é o caso do Tesauro Brasileiro de Ciência da Informação, de autoria de Lena Vânia Ribeiro Pinheiro e Helena Dodd Ferrez.

Baseado em Sales e Café (2009), outra coisa que podemos observar no quadro 1 é que, embora sejam os tesauros e listas de cabeçalhos as linguagens mais utilizadas pelas bibliotecas, existem outras formas de representação do conhecimento, como as ontologias e taxonomias. As ontologias, assim como os tesauros, são linguagens voltadas à organização dos conceitos em determinada área do conhecimento, porém, diferentemente dos tesauros, que trabalham com relações semânticas, as ontologias são organizadas em categorias. Já as taxonomias, segundo Vital e Café (2011), são estruturas organizadas hierarquicamente, do assunto mais específico ao assunto mais genérico ou baseadas em uma relação gênero-espécie.

Cada LD, com suas características distintas, atende a um objetivo específico dentro da documentação, assim, observa-se, no dia a dia das bibliotecas universitárias, que as LDs mais utilizadas para representação dos documentos são os tesauros e cabeçalhos de assunto, já que elas lidam com informações, suportes e meios diversos. Assim, o uso de uma LD está intimamente ligado ao contexto em que a biblioteca está inserida, assim como o público que deseja alcançar. Estas diretrizes são determinadas pela direção da biblioteca ou do sistema de biblioteca e são expressas numa política de indexação, que pode estar formalizada em documento ou apenas subjetivamente adotada pelo consenso entre os bibliotecários.

A política de indexação é importante tanto para as bibliotecas físicas, quanto para os repositórios e bibliotecas digitais, uma vez que elas direcionam e padronizam esta atividade, orientando os profissionais envolvidos quanto a seus procedimentos. Para Nunes (2004, p. 55), uma política de indexação "é uma diretriz que explicita as escolhas técnicas (por isso política) que a biblioteca faz (e os bibliotecários precisam observar em suas rotinas), considerando fundamentalmente duas variáveis: o seu usuário e o seu acervo", isso corrobora com o que foi dito recentemente sobre a escolha da LD mais adequada, considerando o contexto da biblioteca e o público que deseja alcançar.

O estabelecimento e formalização de uma política de indexação é importante para as bibliotecas, pois, segundo Fujita (2016, p. 16), é ela que vai determinar "[...] a exaustividade e a especificidade da indexação, o processo de indexação e a linguagem de indexação no que tange a sua especificidade e exaustividade". Por isso, é necessário não só que a política de indexação exista, mas que esteja sob conhecimento dos profissionais bibliotecários e que estes entendam sua importância e se utilizem dela, visando alcançar o objetivo maior de um repositório ou biblioteca, que é fazer com que a informação contida no documento chegue até o usuário e possa ser por ele utilizada.

Sobre isso, e para demonstrar a importância da indexação, Nunes (2004) baseia-se na quarta lei de Raganathan – poupe o tempo do leitor – discorrendo que a indexação é feita uma única vez, enquanto que o documento pode ser recuperado inúmeras vezes, compreendendo-se, assim, que uma indexação bem executada atenderá mais facilmente à utilização dos documentos. Assim, o processo de tratamento da informação, constitui-se de etapas que têm suas importâncias interligadas, uma leitura superficial pode fazer com que assuntos importantes escapem do indexador, assim como a não utilização de uma linguagem documentária na tradução dos termos pode causar dispersão entre documentos que tratem de um mesmo tema.

Por isso, é necessário não só a presença de instrumentos e diretrizes, mas também o comprometimento da equipe de profissionais que lidam com a indexação, para os quais esta passa, algumas vezes, como uma simples tarefa de "substituir" o título e alguns elementos do texto por palavras retiradas de seu corpo e traduzidas dentro de uma LD. Tal aspecto é notado por Rubi (2012, p. 108), ao mencionar que, na maioria das vezes, a indexação é vista "somente como um processamento técnico que não necessita de procedimentos sistematizados para identificação de assunto, apenas de uma leitura rápida para identificar e extrair os termos para representar o conteúdo". Nota-se, assim, que a visão de uma parte dos profissionais bibliotecários, ainda trabalha a indexação como uma atividade voltada mais para o conteúdo, para sua organização, e menos para o usuário e a recuperação do documento. Desta maneira, trataremos a seguir da representação do conhecimento no ambiente específico dos repositórios institucionais universitários.

3 REPOSITÓRIOS INSTITUCIONAIS UNIVERSITÁRIOS: UM PANORAMA BRASILEIRO

Os objetos informacionais que os repositórios abrigam por meio da interoperabilidade, que é a forma dos sistemas de recuperação se comunicarem entre si, estão disponíveis para serem acessados de qualquer parte do mundo. Assim, os repositórios digitais, por comportarem um grande volume de documentos, necessitam de especial atenção no tocante à indexação de seu conteúdo, para que os objetos não fiquem perdidos em meio à massa informacional e possam chegar com precisão àqueles que dele necessitam.

Sendo os repositórios institucionais das universidades públicas da Região Norte o campo desta pesquisa, é preciso observar e compreender melhor o que são estas ferramentas. Assim, o tópico seguinte tratará dos repositórios e de como se dá sua estruturação.

3.1 O acesso livre e os repositórios digitais

Os repositórios digitais têm tido ampla adesão entre as universidades e os institutos de pesquisa, pois são ferramentas que auxiliam na organização, preservação e disseminação da informação. São úteis não somente para as instituições ou os pesquisadores, mas também para os bibliotecários, que tem ao seu alcance mais uma tecnologia capaz de auxiliá-los no cumprimento, principalmente, da quarta lei de Ranganathan que diz “poupe o tempo do leitor”. Outra lei de Ranganathan que se faz cumprir com a adoção dos repositórios é a quinta lei, que define a biblioteca como um organismo em crescimento, e os repositórios são uma extensão da biblioteca além dos limites físicos, oferecendo informação a qualquer hora, a qualquer pessoa em qualquer canto do planeta que tenha acesso seja a um computador ou a dispositivos móveis (tablets, celulares, etc.). Dessa forma, os repositórios digitais tem sido objeto de estudos de diversos autores como Freitas e Leite (2018), Leite (2009, 2012), Weitzel (2006a, 2006b, 2006c), dentre outros.

Os repositórios digitais nasceram dentro do Acesso Livre (AL) como uma alternativa à divulgação de documentos científicos que até a década de 90 era dominada pelas grandes editoras, mas que já vinha entrando em crise desde a década anterior (MUELLER, 2006). A crise dos periódicos tradicionais se deu em razão, principalmente, dos altos preços cobrados pelas assinaturas e volumes, o que impossibilitava que a maioria das bibliotecas pudessem manter o acesso e suas coleções atualizadas.

Os periódicos eram os meios mais tradicionais de divulgação científica e o alto preço impunha uma barreira no ciclo da ciência, já que os pesquisadores encontravam dificuldades em acessar as pesquisas de seus pares. Assim, surgiu o movimento AL, tendo como principal objetivo democratizar o acesso a essas pesquisas, trazendo o controle da divulgação para mais perto das instituições e dos pesquisadores. Sobre isso, uma ideia presente em Kuramoto (2006) é de que não é justo os pesquisadores terem que pagar para ter acesso aos próprios trabalhos, sendo que as únicas a obterem lucro com as vendas dos periódicos são as grandes editoras; um problema que se agrava quando se trata de pesquisas financiadas com o dinheiro público. Remover essas barreiras de acesso ao conhecimento científico, principalmente as barreiras físicas e financeiras, democratizando o acesso a ele, foi o motor do movimento AL.

A *Budapeste Open Access Intitiative* (BOAI, 2002) diz que por AL entende-se a plena disponibilidade online dos documentos científicos e, para que isso seja alcançado, recomenda que estes sejam disponibilizados através de periódicos eletrônicos de acesso aberto ou através da utilização de ferramentas que permitam que os próprios pesquisadores depositem seus trabalhos, sem o intermédio de uma editora (autoarquivamento).

Essas ferramentas referidas pelo BOAI eram os repositórios, que já existiam desde 1991, quando Paul Ginsparg lançou o ArXiv, um projeto que reunia textos da área de Física para disponibilizá-los entre seus pesquisadores. Sobre o ArXiv o próprio Ginsparg (2011, p. 145), ao comemorar 20 anos daquele que é considerado o primeiro repositório digital do mundo, declarou que “o plano original era de cerca de 100 submissões de artigos em texto completo todos os anos, cada uma armazenada por três meses, até que o sistema de distribuição de papel existente pudesse alcançá-las”. O que se pode compreender aqui é que a ideia inicial não era manter os *pré-prints* (artigos ainda não publicados) permanentemente armazenados, mas sim servir de suporte aos pesquisadores até que eles tivessem acesso aos documentos impressos. Desse modo, os primeiros textos depositados no ArXiv eram restritos à área da Física, sendo ampliado posteriormente para Matemática, ciências não lineares, Ciência da Computação, Estatística, Biologia e Finanças (GINSPARG, 2011).

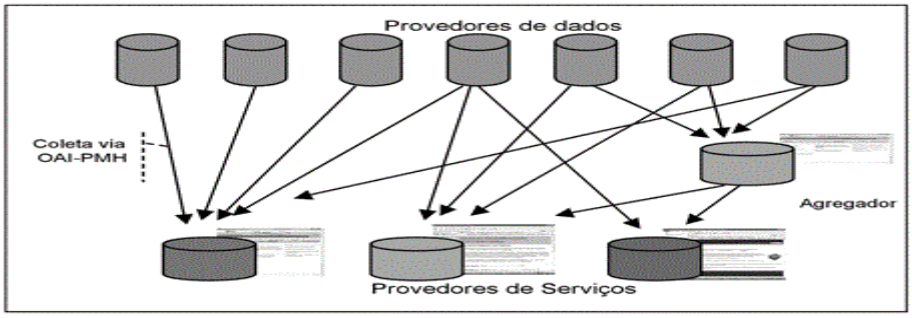
Esses recursos só foram possíveis graças ao aparecimento da *World Wide Web* (www), ou simplesmente *web*, a partir de 1989 (BERNERS-LEE, T. et al., p. 298), o que proporcionou um aumento na divulgação dos textos científicos, utilizando servidores e links (hipertextos) para acessar os documentos. Com isso, a difusão dos textos científicos foi amplamente impulsionada, possibilitando a criação de ferramentas capazes de armazenar e distribuir um grande volume de objetos digitais, como é o caso dos repositórios.

Um repositório, basicamente, constitui-se de uma coleção de objetos informacionais, disponíveis em ambiente digital e online, caracterizando-se pelo uso de metadados para a descrição dos documentos, a licença de distribuição e a disponibilização de conteúdo completo dos textos. Os repositórios digitais são baseados na estrutura de softwares criados com o objetivo de gerir a produção científica em meio digital, sendo o DSpace o mais difundido por permitir o depósito de vários formatos de arquivo (textos, imagens, mapas etc.). O DSpace é um software de código aberto, desenvolvido originalmente pela *MIT Libraries e Hewlett-Packard* (HP) Labs e hoje gerenciado pela DuraSpace.

Para a descrição do conteúdo depositado nos repositórios, o DSpace utiliza metadados padronizados, baseados na iniciativa da *Dublin Core Metadata Initiative* (DCMI). Este padrão é um conjunto composto atualmente de 15 campos para descrição de conteúdo digital, descritos da seguinte forma: *Creator, Contributor, Publisher, Title, Date, Language, Format, Subject, Description, Identifier, Relation, Source, Type, Coverage*, e *Rights*. A adoção de um padrão de metadados, como o Dublin Core, facilita a interoperabilidade na web, pois, estando a descrição padronizada, a integração de informações e recursos disponíveis torna-se mais precisa.

A interoperabilidade já era uma preocupação da *Open Archives Initiative* (OAI) que, segundo Kuramoto (2006), entre suas especificações técnicas, recomendava que os repositórios mantivessem uma *interface* que permitisse a coleta dos metadados por terceiros e, para isso, desenvolveu o *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH). Esse protocolo permite que os sistemas dos repositórios se comuniquem entre si, implicando em uma hierarquização entre eles, dividindo-os em provedores de dados, agregadores e provedores de serviço, como podemos visualizar na figura 1.

Figura 1- Hierarquização entre repositórios



Fonte: Kuramoto (2006, p. 95)

Na Figura 1, pode-se perceber que provedores de dados são as instituições que fazem a coleta dos documentos e os disponibilizam nos repositórios; exemplos de provedores de dados são as universidades federais da Região Norte, como a UFPA e a UFAM.

Os agregadores são instituições que captam os dados disponibilizados pelas instituições provedoras de dados e os disponibilizam em uma interface própria, possibilitando que sejam coletados por outros sistemas maiores. Um exemplo de agregador é a BDTD do IBICT, que reúne a produção inserida nos repositórios e bibliotecas digitais brasileiras.

Os provedores de serviço são sistemas maiores que colhem os dados dos agregadores (que neste caso funcionam como provedores de dados) e os disponibilizam globalmente. Exemplo de provedor de serviço é a *Network Digital Library of Thesis and Dissertation* (NDLTD), desenvolvida e mantida pela *Virginia Polytechnic Institute and State University*.

Ao enxergarmos o fluxo de circulação entre os três níveis e funções de repositórios, podemos compreender a capacidade que eles têm de potencializar o uso da informação, pois um documento que foi disponibilizado numa universidade no norte do Brasil pode ser acessado por um pesquisador residente na Ásia. Isso se dá graças à capacidade de comunicação entre os softwares dos repositórios; o que possibilitou, somando-se à capacidade de disseminação, que os repositórios tivessem rápida aceitação na comunidade científica. Eles são vantajosos tanto para a sociedade que pode se aproximar da literatura científica, quanto para pesquisadores e instituições, que veem sua visibilidade e impacto sendo impulsionados.

Os repositórios, além dessas características mais técnicas, que demonstram seu funcionamento e função dentro das bibliotecas, possuem características que os distinguem segundo a sua natureza, podendo ser temáticos ou institucionais:

- **repositórios temáticos:** como o próprio nome já diz, tratam de um tema ou área específica do conhecimento, como o CarpeDIEN, mantido pelo Instituto de Energia Nuclear (IEN);
- **repositórios institucionais:** são destinados à produção científica de uma instituição, podendo abranger documentos produzidos institucionalmente ou por pesquisadores ligados a ela, como professores e técnicos. É o caso dos repositórios mantidos pelas universidades e institutos de educação e pesquisa.

Sobre esta última categoria, Lynch (2003) caracteriza os repositórios institucionais universitários como “um conjunto de serviços que esta universidade oferece aos membros da

sua comunidade para a gestão e disseminação do material digital, por ela produzido e pelos membros de sua comunidade". Os serviços citados são a gestão, a preservação, a organização e principalmente a disseminação, já que os arquivos abertos aumentam o número de acessos e, consequentemente, de citações, isso faz com que a universidade ganhe maior prestígio diante da comunidade acadêmica (CROW, 2002). O mesmo autor (CROW, 2002, p. 3) também elencou as principais características de um repositório institucional, são elas:

- Institucionalmente definido;
- Acadêmico;
- Cumulativo e perpétuo;
- Aberto e interoperável.

Essas características dos repositórios convergem para o objetivo principal do AL de dar amplo acesso aos documentos científicos e formar uma grande rede informacional. Assim, baseado em Fausto (2013), podemos ver que as primeiras iniciativas em acesso aberto para a divulgação da produção científica exclusiva de uma universidade à nível mundial não foram os repositórios, mas as bibliotecas de teses e dissertações eletrônicas. Isso, talvez, tenha se dado pelo fato de que as teses e dissertações constituíam o maior acervo direto das universidades, já que, na década de 90, os artigos ainda tinham grande vínculo com as grandes editoras.

Mesmo na BOAI, a primeira e principal declaração em favor do AL, os repositórios não são citados, aparecendo apenas uma menção a ferramentas que possibilitassem aos autores depositarem seus trabalhos. Porém ao completar seu décimo aniversário, a BOAI (2012) apareceu munida de novas diretrizes, clarificando as estratégias de depósito, citando diretamente os repositórios como ferramentas de depósito, mostrando preocupação com os direitos autorais (recomendando a utilização da *Creative Commons* como licença de distribuição).

Esse tipo de licença tem por função assegurar ao autor o direito intelectual que ele tem sobre o documento já que o está disponibilizando para amplo acesso na internet, permitindo seu uso sob a condição de que a fonte seja devidamente citada. As licenças permitem ao autor expressar quais tipos de permissões ele está colocando sobre o seu trabalho, por exemplo, se ele permite que, além de acessar, baixar, se faça o uso comercial e a modificação do documento. No caso de repositórios em que os autores delegam a terceiros o depósito de suas pesquisas, essas permissões são expressas em um termo de autorização e autoria (anexo B), que é entregue juntamente com a mídia digital que contém o trabalho.

Assim, o cuidado em aderir a uma licença não só assegura os direitos do autor, como também a credibilidade das universidades que mantém os repositórios.

Com isso, podemos notar que o AL à informação trouxe muitas vantagens para pesquisadores e bibliotecas, assim como desafios a estas últimas, que precisaram desenvolver e se adaptar às novas tecnologias surgidas no seio do AL. Novos instrumentos foram gerados, como os repositórios institucionais que facilitam a comunicação científica, ampliam o uso de documentos, dão maior liberdade aos pesquisadores e melhor controle das pesquisas desenvolvidas pelas universidades. Diante de tudo isto, o tópico seguinte tratará das questões de acesso aberto e dos repositórios no Brasil, o incentivo por parte dos órgãos governamentais responsáveis pela informação, ciência e tecnologia, assim como a adoção destas ferramentas por parte das universidades brasileiras.

3.1.1 Acesso aberto no Brasil e repositórios institucionais brasileiros

O AL ganhava força no mundo todo, e o Brasil não deixou de acompanhar esse movimento, pois pesquisadores, instituições e órgãos governamentais, percebendo a importância de abraçar uma ciência livre, juntaram-se às ações. Surgiram, assim, projetos e incentivos junto às instituições e universidades brasileiras promovendo o uso das ferramentas do AL, como as bibliotecas digitais, as revistas eletrônicas de acesso aberto e os repositórios.

Um marco para o estabelecimento do AL no Brasil foi o lançamento do Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica, assinado e pulicado em 2005. O manifesto mencionava o contexto que possibilitou a formalização do AL no mundo, assim como a importância da informação científica para o desenvolvimento do país. Diante disso, algumas recomendações foram direcionadas visando a aderência ao movimento por parte dos pesquisadores, instituições acadêmicas, agências de fomento e editoras comerciais e não comerciais. Duas recomendações que merecem ser salientadas são a de que as instituições procedessem à criação de repositórios institucionais e temáticos e, ainda, a que conclamava os autores a aderirem às iniciativas, colaborando com o povoamento e a disponibilização de conteúdo (IBICT, 2005).

A partir do manifesto brasileiro, outros grupos no país também se movimentaram para declarar apoio ao movimento AL, assim, podemos destacar a Declaração de Salvador, a Carta de São Paulo, ambas lançadas também em 2005, e a Declaração de Florianópolis, lançada em 2006. Todas as declarações/cartas manifestavam seu entendimento sobre a importância de fortalecer as ações em prol do AL no Brasil, assim como faziam suas devidas

recomendações e chamavam os pesquisadores para colaboração. Desta maneira, essas iniciativas contribuíram imensamente para a disseminação do conceito de AL, assim como para a popularização de suas ferramentas, periódicos eletrônicos de acesso aberto e repositórios.

Muito embora as declarações e cartas supracitadas façam menção ao uso dos repositórios digitais, tanto temáticos, quanto institucionais, as primeiras iniciativas no Brasil para retirar as barreiras de acesso ao conteúdo científico foram as bibliotecas digitais de teses e dissertações. Impulsionado pelas diretrizes do AL e pelo desejo de contribuir com a disponibilização gratuita de textos completos, em 2002, o IBICT lançou a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD).

O objetivo do IBICT, ao criar a BDTD, era de que ela se tornasse uma grande biblioteca geral, que disponibilizaria o conteúdo científico defendido sob a forma de teses e dissertações das universidades brasileiras. Assim, o projeto piloto da BDTD contou, além do próprio IBICT, com a participação de três universidades, sendo elas a Universidade de São Paulo (USP), a Pontifícia Universidade Católica do Rio de Janeiro (PUC-RIO) e a Universidade Federal de Santa Catarina (UFSC). Visando atingir seu objetivo de ser uma agregadora das teses e dissertações em formato digital, o IBICT, juntamente com outras instituições responsáveis, criou e estabeleceu o Padrão Brasileiro de Metadados para Descrição de Teses e Dissertações (MTD-BR), que atualmente está em sua terceira edição, o MTD3-BR, e é baseado no padrão *Dublin Core.* Atualmente, o MTD3-BR é composto, ao todo, por 38 metadados, sendo que, deste total, 18 são metadados obrigatórios e 20 metadados opcionais.

Apesar de o IBICT ser oficialmente o órgão nacional responsável pelo incentivo à informação científica e, consequentemente, o principal difusor das iniciativas em AL no Brasil, paralelamente às suas ações, a CAPES divulgou em 2006 a portaria 013/2006. Esta portaria instituía, em regime de obrigatoriedade, que as teses e dissertações defendidas nos programas de pós-graduações reconhecidos por ela fossem divulgadas digitalmente a partir de março daquele ano. Essa medida, segundo Kuramoto (2010), afetou o crescimento da BDTD, já que a portaria determinava que as universidades depositassem as teses e dissertações em sites próprios ou fornecidos pela CAPES.

As iniciativas dos dois órgãos, apesar de diferentes em alguns aspectos, como o direcionamento para disponibilização dos textos, tinham em comum o seu objetivo principal, que era difundir o amplo acesso a textos científicos no Brasil. Assim, apesar de suas distinções, uma não atrapalhou a outra, como havia citado Kuramoto (2010), graças à

interoperabilidade proporcionada pelos padrões *Dublin Core* e pelo OAI-PMH, a qual permitiu que as duas fossem normalmente executadas sem danos mútuos.

Neste cenário, em 2005, o IBICT apresentou à Financiadora de Estudos e Projetos (FINEP), a proposta de um novo projeto, o Portal de Publicação Seriada de Acesso Livre (PCAL). O PCAL tinha por objetivo atender às recomendações das duas vias do AL: a via amarela, que recomendava o depósito dos artigos em periódicos eletrônicos de acesso aberto, e a via verde, que recomendava a criação de repositórios institucionais (RIs) (KURAMOTO, 2010). Assim, o IBICT buscou englobar em suas ações tanto o software DSpace para desenvolvimento de repositórios, como também um software chamado Sistema Eletrônico de Editoração de Revistas (SEER). Esses dois softwares viriam a compor os pacotes que seriam distribuídos às instituições para difusão das iniciativas em AL.

Mas antes de falarmos mais detidamente desse edital, faz-se necessário citar que o convênio entre o PCAL, a FINEP e a Fundação de Ciências, Aplicações e Tecnologia Espaciais (FUNCATE), constava no termo de compromisso e gestão assinado entre o IBICT e o Ministério da Ciência e Tecnologia (MCT). A verba destinada ao convênio FUNCATE/FINEP/PCAL era de R$ 1.121.839,37 e a construção do Portal de Repositórios e Periódicos Livres constava como um projeto estruturante institucional, a meta era que o portal fosse criado até o final do ano de 2006 (BRASIL..., 2007). O termo visava assegurar ao IBICT condições para o cumprimento de seu Plano Diretor – PDU – 2006-2010, por isso a meta de criação do portal consta até o final de 2006, sendo que o termo só foi assinado em maio de 2007.

O movimento de livre acesso às publicações científicas e de construção de bibliotecas digitais e repositórios ganhava força no Brasil de modo que, em 2009, o projeto FUNCATE/FINEP/PCAL, como já mencionado anteriormente, lançou um edital de chamada pública para que instituições interessadas em implantar repositórios apresentassem seus projetos. As instituições vencedoras foram premiadas com um kit tecnológico que continha o servidor já formatado contendo as plataformas UNIX/LINUX, os softwares básicos Apache, MySQL e PHP, e ainda os softwares DSpace e SEER. O edital FINEP/PCAL/XBDB 002/2009 pode ser visto em sua completude no Anexo B deste trabalho.

Esse edital foi de suma importância para o AL, juntamente com as ações do IBICT, para a implantação de repositórios no Brasil; assim sendo, foram contempladas com o kit tecnológico 32 instituições de pesquisa além do próprio IBICT, que já mantinha seu repositório desde 2004 (MARRA, 2009). Das 32 instituições contempladas, 5 faziam parte

de um projeto piloto do IBICT e 27 foram escolhidas através da seleção regida pelo Edital FINEP/PCAL/XBDB 002/2009, como podemos ver na figura 2:

Figura 2 - Instituições que receberam o kit tecnológico para implantação de repositórios institucionais em 2009.

Fonte: Marra (2014, p. 336)

A partir dessas ações de incentivo, os repositórios institucionais passaram a se popularizar entre as universidades, que viam nessas ferramentas a chance de aumentar sua visibilidade e de seus pesquisadores. Assim, exposto o contexto de entrada dos repositórios institucionais no Brasil, cabe agora entender um pouco mais sobre sua estrutura e funcionamento, o que veremos no tópico a seguir.

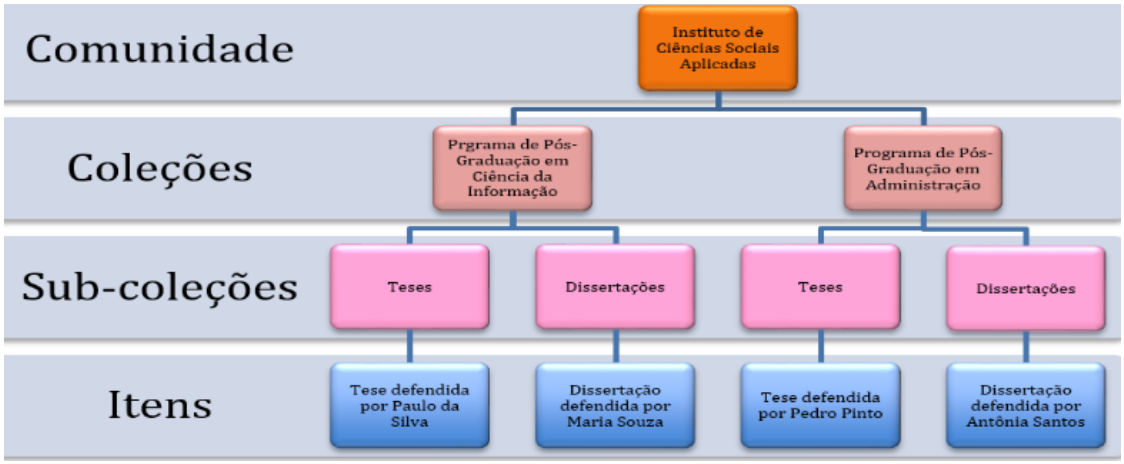
3.2 Estrutura dos repositórios institucionais universitários

Neste tópico trataremos da estruturação dos repositórios institucionais universitários, termo adotado neste trabalho por se tratar de uma pesquisa direcionada aos repositórios mantidos pelas universidades federais da Região Norte do Brasil. Estes repositórios costumam ser chamados, na literatura, apenas de repositórios institucionais, mas entendendo que os repositórios envolvidos neste estudo são mantidos por uma denominação específica de instituição, as universidades, o termo utilizado será repositórios institucionais universitários, assim como visto em Shintaku e Vidotti (2016), Ferreira et al. (2013), entre outros.

Os repositórios institucionais se tornaram importantes ferramentas na gestão da informação dentro das universidades, pois eles não só servem como ferramentas de divulgação, aumentando a visibilidade, mas também são instrumentos de preservação da memória científica. Além disso, os repositórios institucionais universitários também ajudam a sanar um problema recorrente em bibliotecas tradicionais que é a falta de espaço físico para a guarda do volume de informação produzida por alunos, professores e funcionários.

Nos repositórios institucionais universitários, esse volume de informação é organizado de forma hierárquica em: comunidades, coleções e itens. Os itens são os objetos individuais que povoam as coleções e suas subcoleções, que por sua vez compõem as comunidades, que são a estrutura macro dentro de um repositório. No âmbito dos repositórios institucionais universitários, uma comunidade pode ser representada pelas estruturas macros de uma universidade, como é o caso dos institutos (ex.: Instituto de Letras e Comunicação, Instituto de Ciências Sociais Aplicadas, etc.). Sob a coordenação dos institutos estão as faculdades e programas de pós-graduação (ex.: Faculdade de Jornalismo, Faculdade de Biotecnologia, Programa de Pós-Graduação em Ciência da Informação, etc.), estas representam as coleções dentro das comunidades. As coleções podem conter subcoleções (ex.: TCCs, artigos, teses, dissertações, etc.). Os itens são os objetos informacionais que serão inseridos nessas coleções e são a estrutura menor dentro de um repositório, ao qual pode ser mais bem observada na Figura 3:

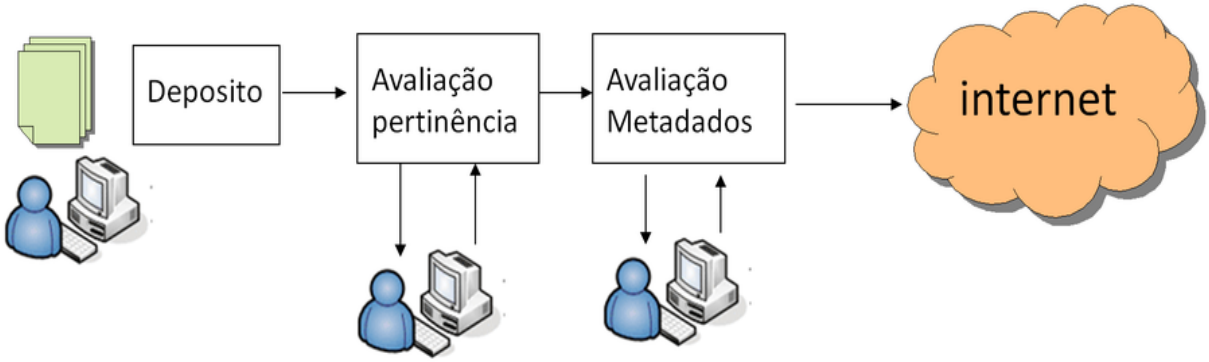
Figura 3 - Estrutura hierárquica dentro dos repositórios institucionais



Fonte: elaborado pela autora, 2020.

Um item, que pode ser uma tese ou dissertação, é depositado pelo próprio autor nas bibliotecas de suas universidades, constituindo, assim, o autoarquivamento, que foi descrito pela *Budapeste Open Access Initiative* (BOAI, 2002) como uma das estratégias do AL, enfatizando que "os estudiosos precisam das ferramentas e da assistência para depositar seus artigos arbitrados nos arquivos eletrônicos abertos". O fluxo de submissão por autoarquivamento, após o depósito, passará ainda por duas etapas: a de avaliação de pertinência e a de avaliação de metadados, o que pode ser observado na Figura 4.

Figura 4 - Etapas de submissão de um item dentro de um repositório.



Fonte: IBICT (2013).

Baseado na figura 4, pode-se afirmar que a avaliação de pertinência dirá se o item corresponde às características das coleções, enquanto a avaliação de metadados se atentará aos campos descritos. Essa etapa é muito importante, pois ela terá forte influência na recuperação dos itens, principalmente na avaliação do campo "assunto", que é um dos que exigem maior atenção, pois ali estão contidos os termos que representarão o conteúdo dos

textos. Portanto, o uso de linguagens documentárias é tão necessário, visto que elas são instrumentos de tradução conceitual, que é, como já dito anteriormente, a passagem dos termos da LN para uma LD, condensando sob uma única denominação preferida para um conceito dentro de determinada terminologia.

Desta forma, diante do entendimento de que o uso de LDs é fundamental para uma recuperação de informação satisfatória e de que isso se intensifica no ambiente virtual, pois o fluxo de documentos disponibilizados é constante e volumoso, será demonstrado a seguir a metodologia utilizada para investigar o fenômeno do uso de LDs nos repositórios institucionais universitários da Região Norte do Brasil.

3.2.1 O campo assunto nos repositórios institucionais universitários

Ao produzir um trabalho acadêmico (teses, dissertações, etc.) o autor lhe atribui palavras-chave, as quais são a forma que o autor dispõe para demonstrar os principais assuntos tratados em sua pesquisa. Porém o autor utiliza-se de sua linguagem natural (mesmo que especializada em determinada área) para fazer a escolha dos termos e, assim, eles podem passar por uma tradução de linguagem natural para linguagem controlada, já que a padronização dos termos resultará numa maior e mais precisa recuperação. Dessa forma, quando o indexador entra em contato com o conteúdo do documento, ele deve pensar tanto nos termos que sintetizam seus assuntos, quanto nos que o tornarão recuperáveis e relevantes aos usuários.

O indexador é caracterizado por Fujita (2004) como um leitor profissional, encarregado de fazer a leitura técnica, utilizando-se de determinados métodos e estratégias para traduzir os textos em termos de linguagem documentária após a análise de seu conteúdo. As estratégias de leitura, como a busca de “pistas” em pontos específicos do texto, possibilitam ao indexador a habilidade de extrair seu assunto principal, traduzindo-o em termos que irão representá-lo. Para o bom funcionamento desse processo, há que existir um bom diálogo entre as variáveis do modelo interativo de Giasson (apud FUJITA, 2004) onde o leitor, o texto e o contexto (psicológico, estrutural, ambiental) dialogam, determinando uma maior ou menor compreensão do assunto do documento. Assim, vê-se que a atividade de indexação não é uma tarefa simples, onde o indexador olha superficialmente para o texto e extrai palavras que julga representá-lo. É necessário que haja conhecimento da estrutura textual, conhecimento prévio ou aprofundado do assunto e instrumentações que permitam a reflexão, especificidade e padronização dos termos.

A adoção de uma ou mais LDs, o nível de especificidade que será tomado no momento da indexação de um documento, serão guiados pela política de indexação estabelecida pela instituição. A política de indexação é um conjunto de diretrizes, destinadas a orientar a atividade de representação temática da informação de forma concisa, baseada em direcionamentos que a padronizam. A política de indexação faz parte de um grupo maior que move a dinâmica de funcionamento de uma biblioteca, orientando de forma específica cada atividade como o atendimento/referência, desenvolvimento do acervo e preservação, etc. (RUBI, 2012).

É necessário que estejam claras na política de indexação as diretrizes a serem seguidas em relação à tradução conceitual (linguagem natural para linguagem controlada) e ao preenchimento do campo "assunto", determinando o nível de exaustividade (alcançar o máximo de termos possíveis dentro do documento) ou especificidade (uso de termos qualificadores até o menor nível de representação). Dessa forma, é necessário fazer com que a informação ali depositada esteja representada da melhor forma possível, caso contrário, há o risco dela se perder em meio à grande quantidade de informação disponível.

Vê-se que, ao mesmo tempo que os repositórios e os recursos *web* proporcionaram muitas melhorias nas atividades tradicionais das bibliotecas, como o problema da falta de espaço para guardar materiais físicos e, principalmente, a disseminação da informação, eles também trouxeram novos desafios. O principal deles é representar os documentos de modo que ele se destaque como informação relevante numa possível busca, fazendo com que ele circule de forma condizente com a missão de fomentar novos conhecimentos. Daí a necessidade de se manter especial atenção com a representação da informação em repositórios, pois, sendo incontável o número de documentos disponíveis e o acesso a eles ocorrendo de maneira remota (sem o auxílio de um bibliotecário), é fundamental que o conteúdo informacional esteja bem representado.

O nível da indexação influencia no que tange à integração de um documento à rede informacional que os repositórios formam, pois isto depende não só da interoperabilidade entre os sistemas e o uso de padrões, mas também dos termos pelos quais estão representados. Por exemplo, um documento que trata de doenças tropicais e tenha sido indexado apenas como "doenças tropicais" terá uma visibilidade reduzida quando uma busca for executada utilizando-se, por exemplo, o termo mais específico "dengue". Por isso, é importante que o indexador se preocupe em representar o assunto do documento não apenas de uma maneira geral, mas de modo que procure conhecer e representar os temas mais específicos e importantes para que o documento possa ser recuperado também por suas facetas menores.

Nos repositórios institucionais universitários, isso implica numa melhor organização dos documentos, já que a indexação que leva em conta apenas o assunto macro implica numa demanda de tempo maior ao pesquisador, que necessitará ler mais atentamente o resumo ou, às vezes, abrir o próprio documento para ter certeza de sua pertinência ou não para sua pesquisa. Nisso, o uso de LDs tem sua importância demonstrada, pois a adoção e uso das linguagens deve levar em conta as particularidades existentes em um repositório institucional universitário, como o fato de englobar diversas áreas do conhecimento e demandar o uso de LDs específicas a elas, não somente uma linguagem geral, como é o caso do *Subject Headings*.

Mesmo sendo um cabeçalho de abrangência geral e, consequentemente, um dos mais ricos e utilizados no mundo, o *Subject* não alcança algumas especificidades demandadas por pesquisas realizadas em áreas como a Amazônia, por exemplo, ou em campos específicos do conhecimento, como a Medicina, que possui vocabulários próprios, mais específicos, como é o caso do *Medical Subject Headings* (MeSH), desenvolvido pela *U.S. National Library of Medicine* (NIH) e o Descritores em Ciências da Saúde (DeCS), desenvolvido pela BIREME.

Assim, se numa biblioteca tradicional o uso de LDs já se faz de extrema importância para a recuperação da informação, isso se torna vital no caso dos repositórios, já que eles estão integrados em rede e seus arquivos são passíveis de acesso e uso em qualquer parte do mundo graças à interoperabilidade proporcionada pelos padrões para descrição de metadados, como já citado anteriormente neste capítulo. No caso dos repositórios que utilizam o padrão *Dublin Core* ou outro padrão baseado nele para a descrição de seus documentos, os assuntos serão descritos no campo "*subject*". O campo "*subject*" não especifica limite máximo de palavras-chave, tampouco permite hierarquização entre termo principal e termo qualificador. Em outras palavras, podem ser inseridas quantas palavras-chave o indexador achar necessário, mas o campo não permite que sejam relacionados separadamente assunto principal e subdivisões, ficando ambos no mesmo espaço, como podemos ver na Figura 5.

Figura 5 - Campo assunto de um item submetido a um repositório.

Fonte: Repositório Institucional da UFPA.

A figura 5 representa a visão do indexador no momento de preenchimento do metadado “assunto” no processo de povoamento em um repositório. Observa-se que não há o estabelecimento de uma hierarquia entre os campos, tendo todos o mesmo valor, já que a indexação para fins de classificação não está presente nos repositórios, fazendo-se presente a indexação para fins de recuperação da informação.

O preenchimento do metadado *subject* proporcionará a organização dos documentos por assunto, que é uma das formas de pesquisa e, consequentemente, recuperação, oferecidas nos repositórios. Para isso, basta o pesquisador abrir o menu de pesquisa e escolher o campo assunto como estratégia de busca, então ele terá acesso a uma listagem alfabética, onde estão dispostos os assuntos sob os quais os documentos estão representados, como podemos ver na Figura 6.

Figura 6 - Tela de navegação pelo índice alfabético de assuntos em um repositório institucional universitário.

Fonte: Repositório Institucional da Fundação Universidade Federal de Rondônia.

Esta tela de navegação demonstra a importância do controle terminológico com termos autorizados através do uso das LDs para sua padronização, pois, caso um repositório adote a linguagem natural, corre o risco de organizar seus documentos de forma dispersa, mesmo que tratem do mesmo assunto. Isto demandaria ao pesquisador tempo para verificar sob qual termo está o documento que melhor lhe convém à pesquisa.

Isto nos permite enxergar ainda mais a importância do uso de LDs para a padronização dos termos utilizados na indexação, visto que isso contribui de forma crucial para a integração dos documentos nos portais de busca, que estão mais distantes dos repositórios provedores de dados. Assim, após esta breve visão sobre o metadado “assunto”, tanto no momento em que o documento está sendo inserido no repositório, quanto no momento em que o pesquisador está executando sua busca, cabe compreender melhor o que são estas ferramentas, o contexto em que surgiram e a sua popularização, assim como o panorama brasileiro de incentivo à sua criação e a sua adoção por órgãos de pesquisa e universidades.

4 METODOLOGIA

Este capítulo é dedicado à metodologia sob a qual se dará esta pesquisa, assim como seu universo e sujeitos. Baseado em Prodanov e Freitas (2013), a pesquisa define-se como uma abordagem qualitativa, observando o fenômeno do uso das LDs por profissionais atuantes nos repositórios institucionais universitários da Região Norte do Brasil. Como estratégia de coleta de dados, será utilizado um questionário semiaberto, abordando questões sobre a atividade dos profissionais que trabalham com a indexação nesses repositórios. Além disso, é de caráter explicativo, caracterizando-se como, primeiramente, uma pesquisa bibliográfica, buscando bases conceituais na literatura existente para entender o que é a OC, a indexação, assim como surgiram e estruturam-se os repositórios institucionais universitários. Caracteriza-se também como um estudo de caso múltiplo, tomando como objeto os procedimentos de representação da informação executados pelos bibliotecários. Para tal, será utilizado como instrumento de coleta de dados um formulário online contendo 23 questões, as quais estão descritas no Apêndice A.

4.1 Descrição do universo da pesquisa e os sujeitos da pesquisa

Esta pesquisa terá como campo de estudo os repositórios institucionais universitários mantidos pelas Universidades Federais da Região Norte do Brasil, esse recorte foi estabelecido por se tratar da região onde a pesquisa nasceu, no caso, dentro da Universidade Federal do Pará.

A Região Norte é composta por sete estados, sendo eles: Acre, Amapá, Amazonas, Pará, Rondônia, Roraima e Tocantins. Segundo consulta realizada no e-MEC, a região possui ao todo dez Universidades Federais, distribuídas de modo que pode ser observado no Quadro 3.

Quadro 3 - Universo da pesquisa

UF DA INSTITUIÇÃO	SIGLA	POSSUI RI?	SITE	SUJEITOS	INSTRUMENTO DE COLETA DE DADOS
Acre	UFAC	Não	--		
Amapá	UNIFAP	Não	--		
Amazonas	UFAM	Sim	http://riu.ufam.edu.br/		
Pará	UFPA	Sim	http://repositorio.ufpa.br/jspui/		

			https://bdm.ufpa.br:8443/jspui/		
			https://livroaberto.ufpa.br/jspui/		
	UFRA	Sim	http://repositorio.ufra.edu.br/jspui/		
	UNIFESSPA	Sim	http://repositorio.unifesspa.edu.br/jspui/		
	UFOPA	Não	--		
Rondônia	UNIR	Sim	http://ri.unir.br/jspui/		
Roraima	UFRR	Sim	http://repositorio.ufrr.br:8080/jspui/		
Tocantins	UFT	Sim	https://repositorio.uft.edu.br/?locale=pt_BR		

Fonte: elaborado pela autora, 2019.

Tendo como base esse escopo, fizeram parte do universo da pesquisa dez repositórios institucionais ativos até o presente ano 2020). Ficaram de fora três das dez universidades federais autorizadas pelo MEC na região Norte e foram incluídos mais dois repositórios pertencentes à Universidade Federal do Pará. Os dados da pesquisa foram obtidos com seis bibliotecários que trabalham com a indexação dos documentos nos repositórios descritos no quadro 3, através do questionário melhor detalhado a seguir.

4.2 Instrumento de coleta de dados: questionário

O instrumento utilizado para coletar os dados será um questionário semiaberto *online* pelo *Google docs*, contendo 23 questões aos profissionais bibliotecários que trabalham como indexadores nos repositórios institucionais universitários da Região Norte do Brasil. As questões foram elaboradas pela autora, com base na literatura consultada e em consonância com os objetivos do estudo e podem ser verificadas no apêndice A.

Os questionários foram enviados a 10 bibliotecas dentro das universidades que compõem o universo desta pesquisa, buscando-se obter uma amostragem de 1 bibliotecário por repositório. Isto possibilitou a obtenção dos dados necessários para a compreensão do uso das LDs na indexação de documentos em repositórios institucionais universitários da

Região Norte do Brasil. As perguntas foram divididas em três seções macro objetivas para melhor divisão das categorias de análise, como destacado no quadro a seguir (4).

Quadro 4 - Distribuição das perguntas e aspectos a serem observados

N. DA SEÇÃO MACRO OBJETIVA	OBJETIVO	CATEGORIAS DE ANÁLISE	QUESTÕES
1	Ambientação do repositório e das atividades dos bibliotecários.	Identificação do repositório.	1 e 2
		Identificação das atividades do indexador.	3, 4 e 5
2	Analisar a adoção ou não adoção de política de indexação e uso de linguagens documentárias na indexação de documentos digitais.	Política de indexação e Linguagens documentárias.	6 e 7
		Uso de Linguagen Natural e Linguagem Documentária (combinada).	8 e 21
		Indexação de documentos digitais.	9
3	Analisar a tradução conceitual de termos para linguagens documentárias.	Análise de assunto	11, 13 e 12
		Tradução conceitual	14 e 15
		Especifidade na indexação.	10, 16 e 17
		Exaustividade na indexação.	18
		Indexação e reuperabilidade.	19 e 20
		Adequação das linguagens documentárias utilizadas e a indexação de documentos científicos na Região norte.	22 e 23

Fonte: elaborado pela autora, 2019.

4.3 Elaboração de categorias de análise

As categorias de análise foram elaboradas de forma conceitual, baseadas em autores da área, como distribuídos no quadro 5.

Quadro 5 - Base conceitual para a elaboração das categorias de análise dos dados

Categorias para análise dos dados	Fonte da categoria
1. Identificação do repositório	Autoria própria a partir da definição do universo da pesquisa.
2. Identificação das atividades dos indexadores	Autoria própria a partir da coleta de dados
3. Política de indexação e linguagens documentárias	Cintra et al. (2002); Fujita e Santos (2016); Rubi (2012), Rubi (2008).
4. Uso de linguagem natural e linguagem documentárias (combinado)	Cintra et al. (2002); Ferreira et al. (2017).
5. Indexação de documentos digitais	Autoria própria a partir do problema de pesquisa.
6. Análise de assunto	Redigolo (2014)
7. Tradução conceitual	Chaumier (1988); Fujita (2003).
8. Especificidade na indexação	Fujita (2015)
9. Indexação e recuperabilidade	Autoria própria a partir da coleta de dados.
10. Adequação das linguagens documentárias utilizadas e a indexação de documentos científicos na Região Norte.	Barité (2007; 2016)

Fonte: elaborado pela própria autora.

Algumas categorias do Quadro 5 foram incluídas após a aplicação do instrumento de pesquisa, pois verificou-se a necessidade de analisar alguns aspectos de forma separada ou forma conjunta dependendo da adequação ao objetivo da pesquisa.

4.4 Forma de análise dos dados

A análise dos dados se deu numa relação 6/1, onde cada categoria acima citada foi observada a partir das respostas de no mínimo 1 e no máximo 4 perguntas. A categorização utilizada na análise dos dados pode ser melhor observada pelo modelo a seguir:

Quadro 6 - Forma que os dados foram analisados

Bibliotecário por ordem de resposta	Questão 1:
1	Resposta
2	Resposta
3	Resposta
4	Resposta
5	Resposta
6	Resposta
7	Resposta
8	Resposta

Fonte: elaborado pela autora, 2019

Para manter a fidelidade das respostas, elas foram transcritas da forma exata como foram recebidas, isto incluiu alguns erros ortográficos e optou-se por identifica-los seguidos da palavra (sic) entre parênteses. A partir disso, foram utilizados gráficos e quadros (o mais adequado para cada situação) para melhor visualização dos aspectos mais importantes identificados nas respostas de cada bibliotecário. Em seguida foi feita análise dos dados obtidos, baseando-se na literatura que orientou cada categoria, para confirmá-las ou refutá-la.

Ao final da exposição das respostas, da montagem dos gráficos e da análise individual das perguntas, dedicou-se um tópico para uma análise conjunta, observando os elos do processo que rege a indexação de documentos nos repositórios digitais universitários da Região Norte do Brasil.

5 ANÁLISE DOS DADOS

Este capítulo foi elaborado após a aplicação dos questionários e o retorno dos respondentes, objetivando-se proceder com a análise detalhada de cada categoria a partir das respostas obtidas. Como descrito na metodologia, foi enviado o questionário para todos os repositórios mantidos por Universidades Federais da Região Norte do Brasil, compreendendo, ao todo, 8 repositórios. Os bibliotecários foram identificados com os números de 1 a 8, de acordo com a ordem do retorno do questionário.

A análise segue a estrutura divisória montada no questionário, ou seja, por seções, categorias e questões. A ordem de análise também segue a estrutura do questionário no tocante às seções e categorias, porém algumas questões foram trocadas de posição, a fim de se alcançar uma melhor compreensão das informações obtidas.

5.1 Seção 1: identificação dos repositórios e das atividades dos indexadores

A seção 1, composta pelas questões numeradas de 1 a 5, teve por objetivo traçar um panorama sobre os RDUF participantes, assim como identificar as atividades desenvolvidas pelos indexadores respondentes da pesquisa. Este subtópico foi dividido em seções menores, a fim de melhor identificar as respostas de cada pergunta pertencente à seção 1.

5.1.1 Identificação dos repositórios participantes da pesquisa

Este subitem é composto pela análise das repostas obtidas para as questões 1 e 2, as quais buscaram identificar os repositórios participantes e o tempo em que cada um está em atividade.

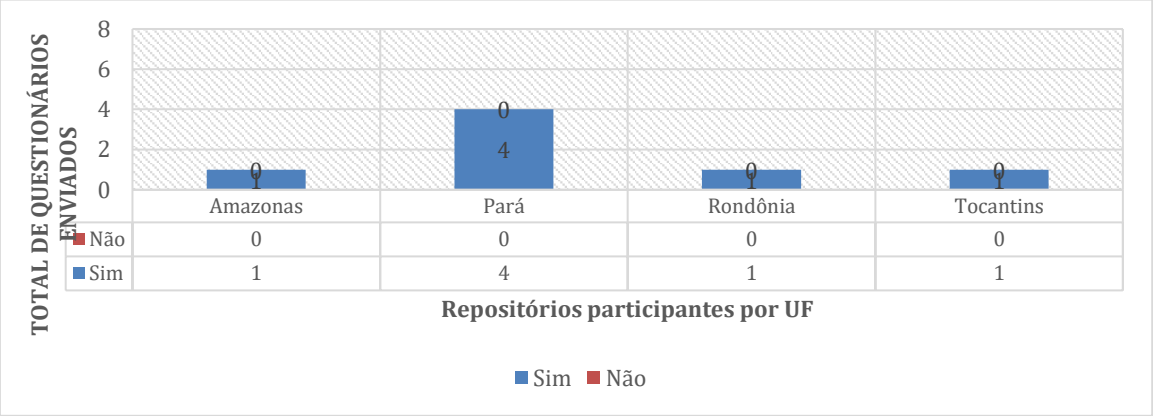
O mapa 1 identifica os repositórios universitários federais existentes na Região Norte, em uma escala de 0 (nenhuma ocorrência) e 5 (maior ocorrência), por unidade federativa, como pode ser observado a seguir:

Mapa 1 - Repositórios por estados da Região Norte

Fonte: elaborado pela autora, 2019.

Como ilustrado no mapa 1, nem todas as Universidades Federais da Região Norte possuem repositórios, sendo elas as dos estados de Roraima, Acre e Amapá. Os estados do Amazonas, Rondônia e Tocantins possuem um repositório cada, enquanto a maior ocorrência se dá no estado do Pará, que possui 5 repositórios federais universitários, distribuídos entre três de suas 4 universidades federais. Dessa forma, estando 5 repositórios localizados no Pará e 3 nos estados do Amazonas, Rondônia e Tocantins, foram enviados um total de 8 formulários. O gráfico 1 ilustra o total de formulários enviados por estado, assim como aqueles dos quais obteve-se as respostas:

Gráfico 1 - Respondentes por estado



Fonte: elaborado pela autora, 2019.

No Gráfico 1, pode-se observar que, dos 8 repositórios que receberam o questionário, nenhum deixou de responder à pesquisa, o que corresponde a 100% de retorno às questões enviadas na pesquisa. Dessa forma, observando a ordem de resposta de cada repositório, estes foram identificados com os números de 1 a 8, sendo esta classificação mais bem identificada no quadro 7, montado a partir das repostas à questão 2, que inicia uma análise mais direta das repostas obtidas individualmente:

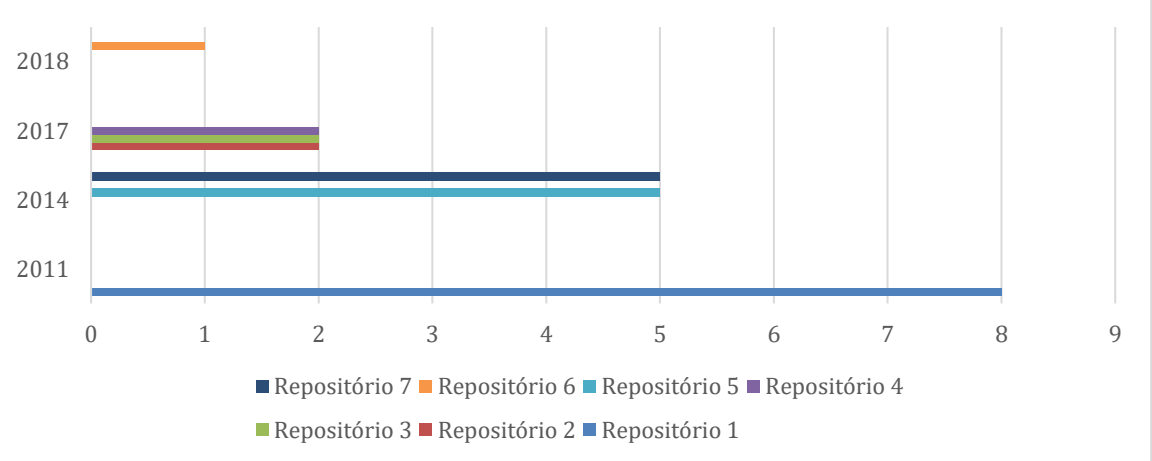
Quadro 7 – ano de início das atividades em cada repositório

Repositório por ordem de resposta	Questão 2: **Ano em que o repositório entrou em funcionamento**
1	2011
2	2017
3	2017
4	2017
5	2018
6	2014
7	2014
8	2019

Fonte: elaborado pela autora, 2020.

A partir desses dados obtidos com a questão 2, foi possível montar o gráfico 2, a fim de melhor demonstrar o ano em que cada repositório entrou em funcionamento:

Gráfico 2 - Ano em que cada repositório entrou em funcionamento.



Fonte: elaborado pela autora, 2019.

O gráfico acima nos mostra que os repositórios começaram a ser implantados pelas Universidades Federais da Região Norte no ano de 2011, sendo mais antigo o repositório identificado com o número 1, estando, portanto, há 8 anos em atividade. Isto ocorreu dois anos anos após o primeiro edital lançado pelo já citado projeto FUNCATE/FINEP/PCAL, que incentivava a criação de repositórios nas universidades brasileiras. Os próximos repositórios a serem implantados iniciaram suas atividades três anos após o primeiro, em 2014, estes foram identificados pela ordem de resposta com os números 6 e 7, estando há 5 anos em funcionamento. No ano de 2017, mais três repositórios entraram em atividade nas Universidades Federais da região Norte, sendo eles identificados pelos números 2, 3 e 4, tendo dois anos de atividade cada um. Os repositórios mais recentes a iniciarem suas atividades foram os identificados pelos números 5 e 8, entrando em atividade nos anos de 2018 e 2019, respectivamente.

Assim, após identificada a quantidade de repositórios universitários federais participantes da pesquisa, bem como seu tempo de funcionamento, pode-se, a partir disso, fazer a identificação das atividades desenvolvidas pelos bibliotecários respondentes, como será descrito no tópico a seguir.

5.1.2 Análise das atividades dos indexadores

Tão importante quanto ter um panorama sobre a existência dos repositórios universitários federais na Região Norte, é também identificar as atividades desenvolvidas pelos bibliotecários que indexam documentos nestas ferramentas. Para isto, foram direcionadas as questões de número 3, 4 e 5, a fim de conhecer se estes profissionais desenvolvem a atividade de indexação com exclusividade ou se ela é desenvolvida concomitantemente a outras atividades do cotidiano bibliotecário.

A questão de número 3 buscou identificar se os bibliotecários desenvolvem apenas atividades relacionadas à indexação de documentos ou se desenvolvem também outras atividades. O quadro 8 detalha as repostas dadas pelos indexadores, obedecendo à ordem na qual as respostas foram recebidas.

Quadro 8 - Atividades desenvolvidas pelos indexadores.

Bibliotecário por ordem de resposta	Questão 3**: Você trabalha exclusivamente com indexação de documentos ou desenvolve outras atividades?**
1	"Indexação de documentos e outras atividades"
2	"Indexação de documentos e outras atividades"
3	"Não, faço a revisão da área de conhecimento do FICAT 2.0"
4	"Outras atividade (*sic*) como validação de informações na BDTD e RIU, Organização de Material impresso, Contato com autores sobre trabalhos confidenciais, auxilio a usuários, dentre outros."
5	"Gerencio as atividades do RIUFRA e também os sistemas de informação da biblioteca"
6	"Gerência"
7	"Não, planejamento e gestão do RIUFT"
8	"Não"

Fonte: elaborado pela autora, 2020.

O Quadro 7 demonstra que a totalidade dos bibliotecários, ou seja, 100% dos respondentes não desenvolvem somente atividades relacionadas à indexação de documentos, sendo esta conciliada com outras funções, como a de gerência de sistemas de informação e atividades administrativas relacionadas ou não com o repositório. Isso indica que não há um direcionamento exclusivo de profissionais para a representação da informação e que estes costumam desenvolver múltiplas funções no seu fazer profissional. Assim, sabendo que as universidades disponibilizam pesquisas de diversas áreas do conhecimento, a questão de número 4 buscou identificar se os bibliotecários, além de desenvolverem outras funções dentro do sistema, trabalham com a indexação de documentos em áreas específicas ou gerais do conhecimento. Diante desta questão, foram obtidas as respostas identificadas no Quadro 9.

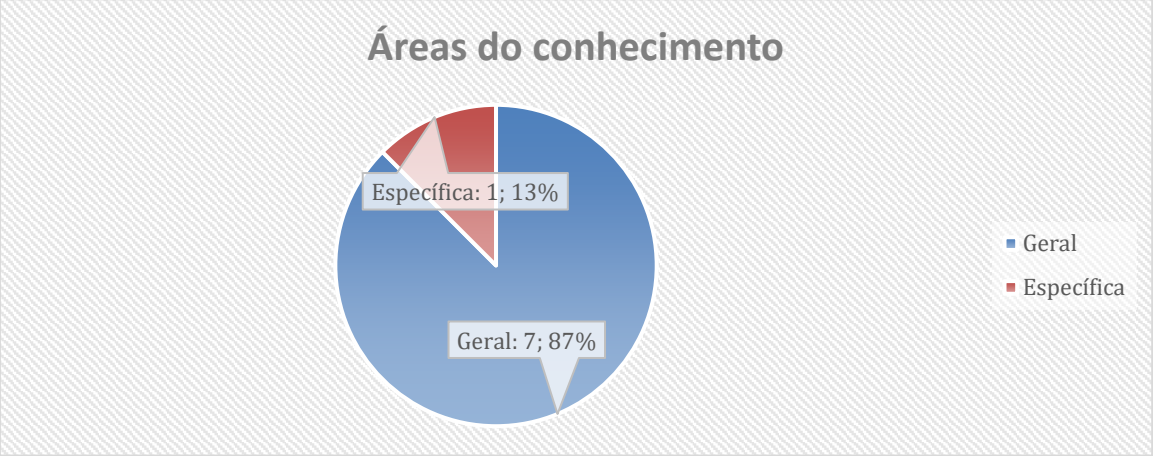
Quadro 9 - Identificação das respostas obtidas na questão 4

Bibliotecário por ordem de resposta	Questão 4: **Você trabalha com a indexação de documentos de diversas áreas do conhecimento ou de uma área específica?**
1	“Sim”
2	“Sim”
3	“Sim”
4	“Sim”
5	“Sim”
6	“Sim”
7	“Não”
8	“Não”

Fonte: elaborado pela autora, 2020.

Para a visualização dos dados obtidos, foi elaborado o Gráfico 3, de acordo com cada resposta recebida.

Gráfico 3 - Áreas do conhecimento onde atuam os bibliotecários.



Fonte: elaborado pela autora, 2020.

Através do Gráfico 3, pode-se notar que a grande maioria dos bibliotecários respondentes (7) trabalha com a indexação de documentos em áreas gerais do conhecimento, o que corresponde a 87% do total. Apenas um respondente trabalha com áreas específicas do conhecimento, o que representa 17% do total. Esse panorama sugere que a maioria dos bibliotecários, caso utilize LDs, adota vocabulários mais gerais, como a BN e a LCSH, porém isso será investigado mais à frente, confirmando ou não essa primeira impressão.

No prosseguimento da pesquisa, a questão de número 5 buscou complementar a questão de número 4, identificando em quais áreas os indexadores desenvolviam suas

atividades, caso fosse em uma área específica. As respostas dos bibliotecários à questão 5 estão descritas abaixo, no Quadro 10:

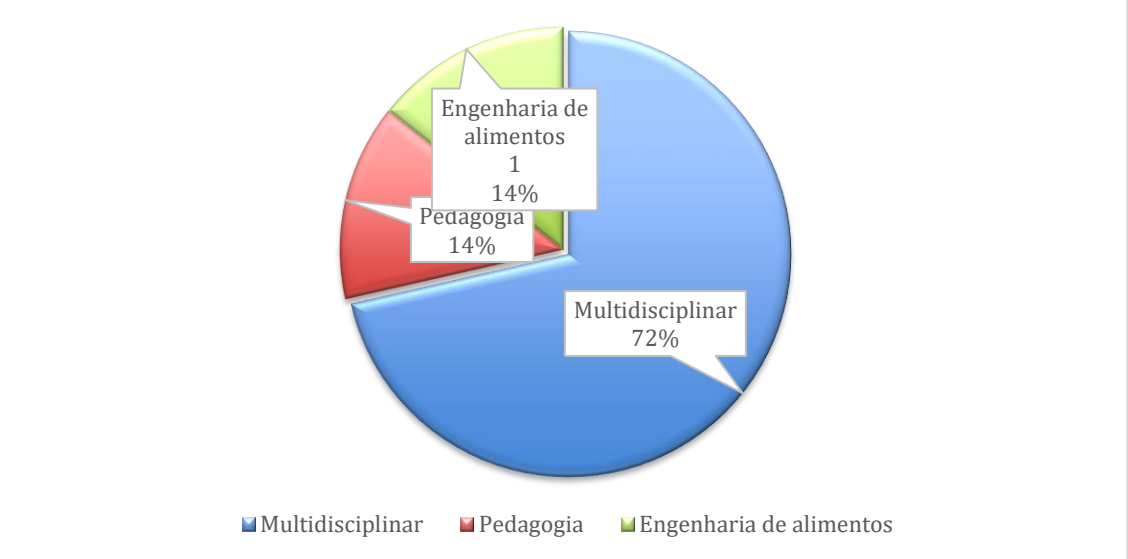
Quadro 10 – Especificidade das áreas do conhecimento onde atuam os bibliotecários.

Bibliotecário por ordem de resposta	Questão 5: **Caso seja específica, em qual área você trabalha? Costuma sentir dificuldade na tradução dos termos para uma linguagem documentária?**
1	"Multidisciplinar"
2	"Multidisciplinar"
3	"Não se aplica"
4	"Não é específica"
5	"Não é específica, é geral"
6	"Trabalho especificamente com a alimentação documentação para os cursos de Pedagogia e Engenharia de Alimentos. Sinto dificuldade com a sub áreas destinadas ao de Engenharia de Alimentos"

Fonte: elaborado pela autora, 2020.

Diante das respostas transcritas no quadro 10, a fim de melhor visualizar os resultados coletados, elaborou-se o Gráfico 4, disposto a seguir:

Gráfico 4 - Atuação dos bibliotecários por área do conhecimento



Fonte: elaborado pela autora, 2020.

As repostas à questão 5 demonstram que os bibliotecários, além de desempenharem outras funções além da indexação (questão 3), também costumam desenvolver esta atividade dentro da multidisciplinaridade nas áreas do conhecimento. Apenas outras duas áreas foram

mencionadas como campo de atuação específica, cada uma abrangendo 14% no total de citação; enquanto a área “Multidisciplinar” foi citada por 74¨% dos respondentes. Houve apenas uma menção quanto a dificuldades na indexação de documentos das áreas específicas, sendo os termos qualificadores (subdivisão de assunto) o ponto citado como a razão das dificuldades na área de Engenharia de Alimentos. Essa questão se reflete na especificidade da indexação, já que os termos qualificadores (subdivisões) são essenciais para o aprofundamento dos assuntos tratados no documento.

Propondo uma síntese sobre os aspectos observados diante das respostas sobre as atividades dos indexadores, percebe-se que a indexação não é um campo exclusivo de atuação entre a maioria desses profissionais. Percebe-se também que a área multidisciplinar tem prevalência sobre as áreas específicas dentro das atividades dos bibliotecários, exigindo, assim, que eles tenham percepção múltipla sobre a diversidade do conhecimento.

Assim, após a obtenção de dados sobre os repositórios e sobre a atuação dos indexadores, a pesquisa segue para pontos mais exatos sobre a indexação de documentos nos repositórios, como a existência ou não de política formalizada que guie esta atividade.

5.2 Seção 2: política de indexação e uso das linguagens documentárias

A seção 2, composta pelas questões numeradas de 6 a 10, objetivou identificar aspectos sobre uma política que sirva de diretriz para a atividade da indexação nas Universidades Federais da Região Norte. Foram abordados pontos como a existência da política de indexação, a indexação de documentos digitais, as recomendações sobre o uso de linguagens documentárias e de linguagem natural. Os dados obtidos por meio das questões 6, 7, 8, 9 e 10 foram detalhados nas subseções a seguir.

5.2.1 Adoção de política de indexação

A questão de número 6 objetivou diagnosticar a existência de uma política de indexação nas Universidades Federais do Norte do país, pois as diretrizes que a política recomenda são importantes para guiar os procedimentos dos bibliotecários dentro da indexação e tornar o catálogo mais uniforme. O Quadro 10 demonstra as respostas obtidas, as quais serviram de base para a elaboração do diagnóstico.

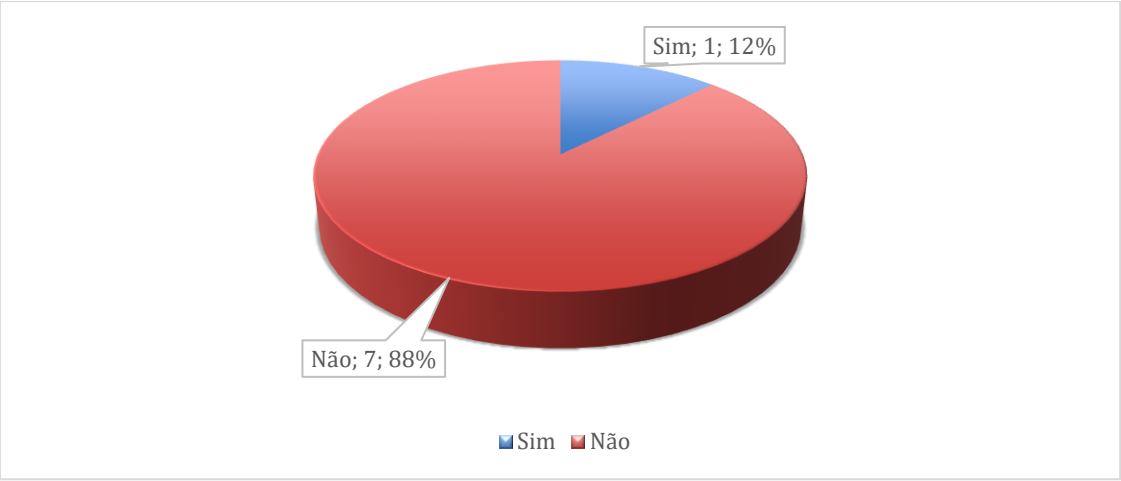
Quadro 11 - Existência da política de indexação nas bibliotecas mantenedoras dos repositórios.

Bibliotecário por ordem de resposta	Questão 6: **Existe em sua instituição uma política de indexação que determine o uso de linguagens documentárias?**
1	"Não"
2	"Não"
3	"Não"
4	"Sim"
5	"Não"
6	"Não"
7	"Não"
8	"Não"

Fonte: elaborado pela autora, 2020.

O quadro acima demonstra que a maioria das instituições não possui uma política que direcione a atividade da indexação, consequentemente, isso se estende aos repositórios digitais. Para melhor visualizar o quantitativo das repostas para "sim" e para "não", elaborou-se o gráfico 5.

Gráfico 5 - Quantitativo das universidades federais nortistas que possuem política de indexação



Fonte: elaborado pela autora, 2020.

Nota-se no Gráfico 5 a prevalência das universidades federais nortistas que não possuem uma política direcionada à indexação de documentos, 7 delas não possuem a diretriz, o que representa 88% da amostra da pesquisa. Assim, somente uma das instituições

respondentes possui uma política de indexação que parametrize essa atividade, o percentual que ela representa corresponde a 17% do total.

A política de indexação, como assinala Rubi (2008), têm a importante característica de padronizar as atividades dos indexadores, diminuindo as subjetividades, já que os profissionais são diferentes entre si, podendo, assim, essas diferenças serem refletidas na construção do catálogo. Fujita e Santos (2016) reforçam essa concepção ao afirmar que a subjetividade advinda das individualidades dos indexadores reflete na recuperação da informação disponibilizada pela biblioteca e, por isso, as instituições devem optar pela adoção de uma política que sirva de diretriz para suas atividades, incluindo neste caso a indexação de documentos.

Um dos principais componentes da política de indexação são os procedimentos quanto à escolha dos termos que representarão o documento, se a linguagem utilizada deve ser unicamente controlada, natural ou situações onde ambas podem ser adotadas. Assim, o tópico seguinte buscou identificar o tipo de linguagem adotada pelos indexadores nos repositórios participantes da pesquisa.

5.2.2 Uso de linguagem documentária, linguagem natural ou ambas

O tópico anterior mostrou que a maioria das bibliotecas das universidades federais do norte do país não possui uma política que rege a indexação de seus documentos, ainda, apenas uma delas possui tal diretriz. Desta forma, as questões de 7 a 10 foram elaboradas de modo a observar as estratégias utilizadas pelos bibliotecários no momento de indexar os documentos a serem inseridos no repositório. Assim, cabe analisá-las separadamente para melhor identificar os dados obtidos.

A questão de número 7 buscou identificar a adoção das linguagens documentárias pelos indexadores, seja por meio e diretrizes presentes em uma política, seja por estratégias pessoais. O quadro 12 identifica as respostas dadas pelos bibliotecários:

Quadro 12 - Tipo de linguagem adotada no repositório

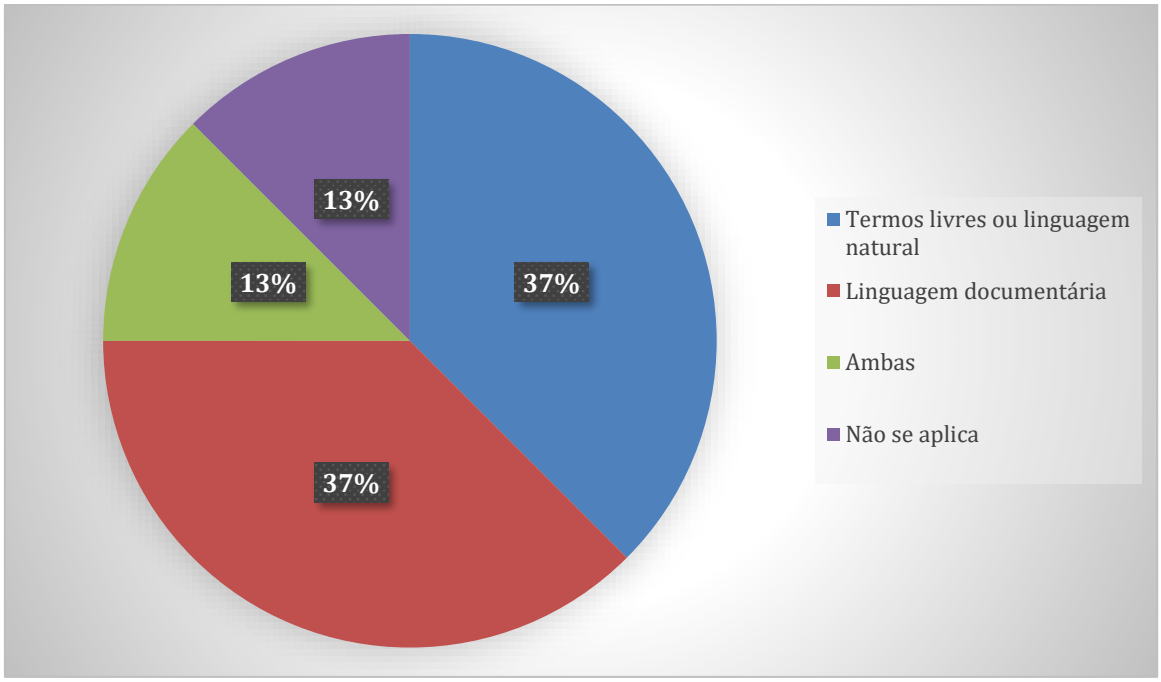
Bibliotecário por ordem de resposta	Questão 7: **Caso a resposta da pergunta 6 seja SIM, sobre o uso de linguagens documentárias, quantas e quais linguagens ela recomenda? Caso seja NÃO, quantas e quais você costuma utilizar e por quê?**
1	"Utilizamos termos livres conforme se apresenta no trabalho e optamos por acrescentar as -chave, os assuntos da ficha e os assuntos do Pergamum"
2	"Não se aplica"
3	"Utilizamos termos livres, conforme se apresenta no trabalho"
4	"Normalmente a BN, *Library of Congress*, BVS, GBV..."
5	"Usamos a linguagem natural, pois a política de indexação ainda está em discussão"
6	"Normalmente adoto a Classificação Decimal Universal para sintetizar os assuntos indexados."
7	"Termo tópico – cobertura de assunto (central e periférico)"
8	"Não usamos"

Fonte: elaborado pela autora, 2020.

As respostas obtidas na questão 7 nos dão a visão de alguns aspectos especiais presentes na atividade de indexação de documentos nos repositórios universitários federais da Região Norte, como a prevalência da utilização de linguagem natural ou termos livres. Para uma melhor observação destes aspectos, foram elaborados gráficos separados a fim de demonstrar cada um deles.

No Gráfico 6, foi feito um comparativo entre o uso das linguagens, natural e documentária, de acordo com as informações contidas nas repostas:

Gráfico 6 - Tipo de linguagem de indexação utilizada nos repositórios.



Fonte: elaborado pela autora, 2020.

O que pode ser observado no Gráfico 6 é que os indexadores, diante da ausência de uma política de indexação, costumam utilizar a linguagem natural para representar os assuntos mais importantes contidos no documento. Os indexadores que optam pelo uso da linguagem natural representam 37% do total de respondentes. Este é um número considerável diante da importância já demonstrada das linguagens documentárias para que se evite a dispersão documentária, o que pode ocorrer através da sinonímia, da grafia diversa de determinado termo e das variações que este tem nas várias línguas e culturas (Cintra et al., 2002). Se voltarmos ao quadro 11, notaremos que das bibliotecas que disseram utilizar a estrutura de LDs (37% do total), uma delas é a que possui política de indexação. Este aspecto demonstra que a política de indexação é importante no que tange a direcionar as atividades do indexador, propondo que ele tome o cuidado de fazer a tradução dos termos escolhidos, como reforçam Ferreira et al. (2017) ao recomendarem que o indexador busque verificar nas LDs os termos autorizados, como garantia de uma representação precisa e padronizada. A precisão e a padronização são elementos muito importantes para a recuperação da informação no momento da pesquisa. A precisão evita que o pesquisador seja forçado a "folhear" o documento para saber de que ele trata exatamente, já a padronização atua para que documentos diferentes sobre um mesmo assunto fiquem dispersos, cada um indexado de uma forma, daí a importância de adoção de uma LD.

Assim, sobre os indexadores que afirmaram fazer uso de uma LD, identificou-se que um deles optou por utilizar uma linguagem enciclopédica, a Classificação Decimal Universal (CDU), para sintetizar os assuntos contidos no documento. Já o respondente que assegurou adotar LD alfabética costuma utilizar-se tanto de linguagens gerais, como o Catálogo de Autoridades da BN e o LCSH, como também de linguagens específicas das áreas de conhecimento, tais como o Descritores em Ciências da Saúde (DeCS), vocabulário controlado mantido pela Biblioteca Virtual em Saúde para a indexação de seus documentos.

A questão 8 foi elaborada de modo a completar a questão anterior, identificando o uso combinado da LN com a LD e verificando se isso estava descrito em política formalizada. Como percebeu-se na questão 7, a maioria dos repositórios não possui uma política de indexação elaborada e dão preferência à utilização da LN. As repostas obtidas diante da questão 8 estão descritas no Quadro 13.

Quadro 13 – Estratégia do uso de linguagens de indexação nos repositórios

Bibliotecário por ordem de resposta	Questão 8: **Ainda sobre política de indexação, caso exista em sua instituição, ela permite o uso da linguagem natural juntamente com a linguagem documentária? Quais os procedimentos?**
1	"Não se aplica"
2	"Não se aplica"
3	"Não se aplica"
4	"Sim. Na política, a linguagem natural pode entrar como remissiva. Na BDTD e RIU, a linguagem natural é normalmente sugerida pelo autor em suas palavras-chave."
5	"Só usamos linguagem natural de acordo com os assuntos do documento"
6	"Não."
7	"Estamos em fase de discussão e elaboração da política de indexação da UFT – temos uma comissão de bibliotecários (5) elaborando uma minuta para apresentação na reunião semestral do Comitê gestor do Sisbib para fechamento da proposta pelo conjuntos (*sic*) dos bibliotecários e apresentação a gestão superior e aprovação no Consuni."
8	"Somente linguagem natural, por enquanto"

Fonte: elaborado pela autora, 2020.

Como citado anteriormente a respeito da questão 8, ela foi elaborada com objetivo de complementar a informação buscada na questão 7. A questão 7 já havia identificado que a maioria dos repositórios não possui política de indexação e opta por utilizar a LN, respeitando os termos utilizados pelos autores, ou uma combinação da LN com a LD. Isto, portanto, demonstra que as repostas obtidas no quadro 12 acima só reforçam a ideia de que quando não existe uma política de indexação, não existe uma recomendação formalizada quanto aos procedimentos a serem adotados na representação da informação. Assim, não há um registro que parametrize a forma como a informação deve ser tratada e como esta deve ser inserida nos sistemas de recuperação.

Sobre o uso de LN ou LD ou ambas, a única política de indexação identificada determina que, em respeito à particularidade de termos regionais, sejam adotadas linguagens nacionais e, se ainda assim o termo pesquisado permaneça sem resposta, os termos utilizados pelos autores devem ser inseridos no registro de autoridades da instituição, tornando-se, assim, um termo autorizado naquele sistema (FERREIRA et al, 2017).

Diante deste panorama acima identificado, observou-se que seria necessário investigar se os indexadores adotam um cuidado especial quanto à indexação do documento digital, visto que esse documento é localizável em uma rede informacional através de acesso online e remoto. Assim, a próxima categoria de análise buscou identificar se esse cuidado especial existe e se está registrado.

5.2.3 Indexação de documentos digitais

Os documentos digitais são disponibilizados de forma online, possibilitando um acesso remoto, ou seja, o pesquisador pode acessar em qualquer parte do mundo, tendo contato direto com a pesquisa e os documentos que ela retorna, sem a interferência de um bibliotecário. A interferência que o bibliotecário é capaz de fazer, diante desta particularidade do acesso remoto, é na hora de indexar o documento e inseri-lo no sistema, tomando cuidado com os termos escolhidos para sua representação.

Desta forma, o quadro 14 foi elaborado de modo a identificar as respostas obtidas através da questão 9.

Quadro 14 - Indexação de documentos digitais

Bibliotecário por ordem de resposta	Questão 9: **Existem recomendações específicas para a indexação de documentos digitais? Se sim, qual os aspectos e cuidados principais que ela recomenda?**
1	"Utilizamos termos livres conforme se apresenta no trabalho e optamos por acrescentar as palavras-chave, os assuntos da ficha e os assuntos do Pergamum"
2	"Utilizamos termos livres conforme se apresenta no trabalho"
3	"Não"
4	"Desconheço"
5	"Não existe"
6	"Não"
7	"Sim, termo tópico com cobertura de assuntos"
8	"Não"

Fonte: elaborado pela autora, 2020.

Percebe-se, pelas repostas encontradas no Quadro 14, que não existe um procedimento adotado em relação às particularidades dos documentos e sistemas de recuperação digitais, adotando-se as mesmas estratégias utilizadas ao indexar o acervo físico, de acesso presencial.

Depreende-se também daí que há uma preocupação por parte das universidades em disponibilizar o conteúdo de sua produção científica para acesso remoto, porém, parece não haver a mesma percepção em relação aos documentos digitais, que necessitam de especial atenção na hora de serem disponibilizados online. Um dos pontos que reforçam esta necessidade é o volume informacional disponível, nacional e internacionalmente, o que demanda uma indexação mais precisa, pois há o risco do documento se perder no amontoado de informação que é disponibilizado todos os dias na internet.

Outro aspecto importante a ser observado é a quantidade de termos utilizados pelos indexadores para representar o assunto dos documentos, pois deve-se pensar na recuperação de toda e qualquer informação relevante para uso da clientela do sistema de informação. Para o conhecimento de tal aspecto, a questão 10 abordou a determinação mínima de termos pra a indexação dos documentos e as respostas recebidas foram transcritas no quadro 15.

Quadro 15 - Média de termos utilizados pelos indexadores.

Bibliotecário por ordem de resposta	Questão 10: **A política de indexação, caso exista na instituição, determina uma quantidade de termos a serem utilizados na representação dos documentos? Se SIM, qual a média ela recomenda? Se NÃO, qual a média que você costuma utilizar?**
1	"No mínimo 1 termo"
2	"No mínimo um termo"
3	"Não. No mínimo 1 termo"
4	"Recomenda-se pelo menos 4 termos."
5	"Não. Utilizamos uma média de 5 ou 6 termos."
6	"Não, adoto em média três."
7	"Não, até 4 termos tópicos"
8	"Não tem política de indexação, por enquanto."

Fonte: elaborado pela autora, 2020.

Observa-se então que a média de termos tópicos utilizados varia entre um mínimo de 1 a um máximo de 6 termos. Essa variação indica que, na ausência de uma política de indexação, os bibliotecários utilizam do bom senso diante de cada documento para determinar quantos termos serão representados para que o conteúdo documental seja resumido e indexado.

Fazendo-se um apanhado das informações dispostas neste subitem, verificou-se que:

a) a adoção de uma política de indexação não é unânime entre as universidades federais nortistas, existindo apenas uma instituição da região que possui tal diretriz;

b) no caso das instituições que não têm uma política de indexação formalizada, é dada preferência à utilização da LN;

c) os repositórios que utilizam LD ainda são minoria, existindo apenas dois que adotam preferencialmente a tradução conceitual para termos controlados por linguagens documentárias ou que utilizam tanto a LN quanto a LD concomitantemente.

Deste modo, percebidos alguns aspectos sobre o uso de LDs nos repositórios, partiu-se para a análise das questões que compõem a terceira e última seção do questionário, a qual engloba os procedimentos sobre a tradução conceitual e aspectos fundamentais da indexação, como especificidade e exaustividade.

5.3 Seção 3: Tradução conceitual dos termos de indexação para linguagens documentárias

A tradução conceitual é um elemento importante na atividade de indexação, pois ela permite escolher o termo que melhor se adequa para representar um documento dentro de uma LD. Essa tradução conceitual permite a padronização dos termos indexadores, evitando a dispersão documental, a sinonímia e as variações gramaticais presentes na linguagem natural. Diante disto, as questões componentes da seção 3 buscaram coletar informações referentes aos procedimentos adotados na etapa de tradução conceitual, quando esta está presente na indexação. Desta forma, as categorias de análise que compõem a seção 3, assim como as questões direcionadas a investigá-las, estão descritas nos subtópicos a seguir.

5.3.1 Análise de assunto

Esta categoria é composta pelas questões 11, 12 e 13, direcionada a identificar os procedimentos adotados durante a análise dos assuntos relevantes a serem indexados no documento. Assim, primeiramente, a questão de número 11 buscou identificar se era feita uma tradução conceitual ou se eram utilizados os termos já previamente adotados pelo autor, ao identificar as palavras-chaves na apresentação do resumo. O quadro 15 identifica as repostas obtidas com cada bibliotecário, na ordem em que os questionários foram respondidos.

Quadro 16 - Análise de assunto.

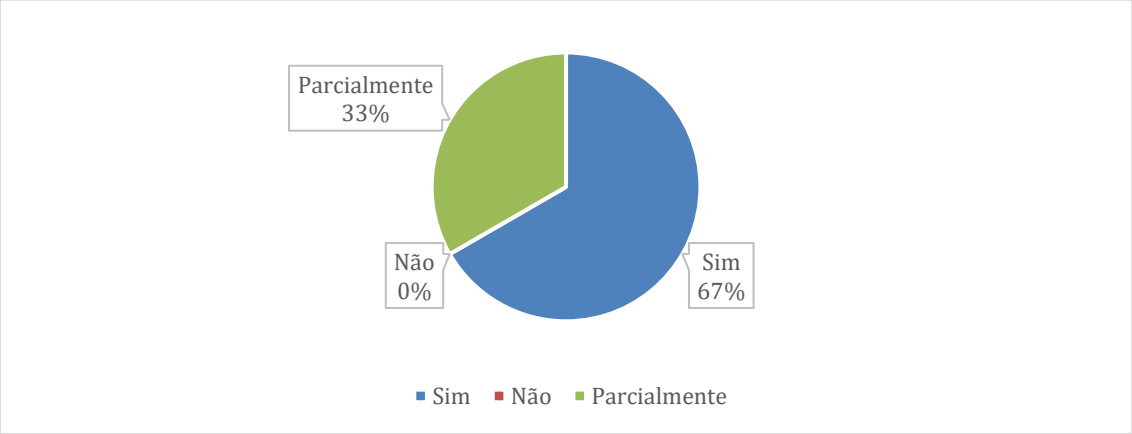
Bibliotecário por ordem de resposta	Questão 11: **Você costuma levar em conta os termos utilizados pelos autores como palavras-chave no momento da representação documentária? Justifique.**
1	“Sim. Respeito a escolha do autor q teve orientação de seu orientador e passou por uma banca especializada”
2	“Sim, considero a expertise do autor”
3	“Sim. Em respeito a escolha do autor q teve orientação de seu orientador e passou por uma banca especializada”

4	“Sim. Normalmente os deles sempre ficam e acrescentamos os termos autorizados.”
5	“Alguns sim, mas não todos, excluímos os termos muito gerais ou que não condizem muito com os assuntos do documento.”
6	“Sim, observo as palavras chaves quando tenho dificuldade de especificar o assunto. O autor do trabalho tem familiaridade com o tema e por vezes consegue dar um norte quanto a melhor linguagem a ser utilizada.”
7	“Sim, pois os autores dentro de suas áreas conhecem os termos gerais e específicos dos seus objetos de estudos”
8	“Termos autorizados pela BN”

Fonte: elaborado pela autora, 2020.

A partir das repostas apresentadas no Quadro 15, é possível fazer uma análise mais aprofundada sobre a utilização, por parte dos indexadores, dos termos apresentados pelo autor ao final do resumo. Desta forma, o gráfico 7 demonstra a média de utilização destes termos no momento da indexação.

Gráfico 7 - Adoção das palavras-chaves utilizadas pelo autor



Fonte: elaborado pela autora, 2020.

A principal informação que se pode visualizar no Gráfico 7 é a de que a maioria dos bibliotecários costuma adotar as palavras-chaves indicadas pelo autor do documento. Assim, esta informação complementa o que foi observado com a questão de número 9 da seção 2, na

qual observou-se que a maioria dos bibliotecários adota a LN na hora de indexar um documento no repositório. Como já citado ao final da seção 2, é necessário ter atenção especial com a indexação de documentos digitais, pois a sua recuperação remota não conta com a participação do bibliotecário nas estratégias de busca do pesquisador. O uso da LN permite o aparecimento da sinonímia e, consequentemente, ocasiona uma dispersão documentária, assim como uma recuperação insatisfatória. A simples adoção dos termos inseridos pelo autor como palavras-chaves possibilita também que a indexação se torne superficial e não represente com exatidão os assuntos do documento no sentido de sua recuperação. Desta forma, fez-se necessário investigar se os bibliotecários executam a análise de assunto e quais partes eles percorrem para a escolha dos termos indexadores. Foi com este objetivo que a questão 13 foi elaborada e as respostas recebidas foram transcritas no quadro 17.

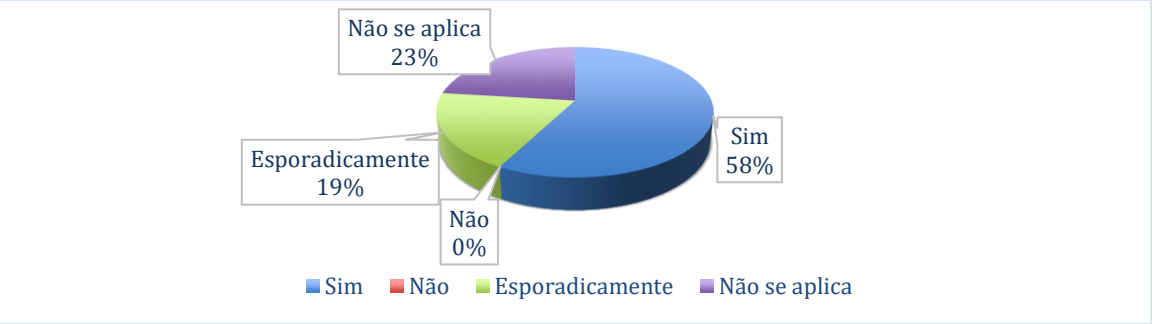
Quadro 17 - Execução da análise documentária por parte dos bibliotecários

Bibliotecário por ordem de resposta	Questão 13: **A tradução dos termos ocorre a partir da análise de assunto (leitura, identificação e seleção dos conceitos)?**
1	“Não se aplica”
2	“Não se aplica”
3	“Em alguns casos, sim”
4	“Sim.”
5	“Sim.”
6	“Sim.”
7	“Sim”
8	“Sim”

Fonte: elaborado pela autora, 2020.

Para melhor análise dos elementos contidos nas respostas do quadro acima, foi elaborado o gráfico (7), inserido abaixo:

Gráfico 8 - Análise documentária por parte dos bibliotecários



Fonte: elaborado pela autora, 2020.

A leitura técnica, segundo Dias e Naves (2013), é uma leitura executada pelo bibliotecário a fim de que este conheça os assuntos do documento e possa, assim, representá-lo. Ainda segundo Dias e Naves (2013), esta leitura é executada seguindo o percurso de determinadas partes do documento, como, por exemplo, título, resumo, sumário e introdução. Assim, diante destas concepções e das respostas obtidas junto aos bibliotecários participantes da pesquisa, percebe-se que a importância do processo é entendida e adotada, mesmo diante da ausência de uma política de indexação como diretriz e suas recomendações referentes à tradução de termos para uma LD.

A leitura técnica é a base da análise de assunto e é essencial para que não se adote apenas termos superficiais que aparecem no documento. A percepção individual do bibliotecário em relação a isso e ao papel que uma boa representação tem na recuperação da informação é fundamental para a satisfação do usuário com o catálogo. Porém, assim como constatado por Redigolo (2010; 2014), muitos indexares não executam a análise de assunto, influenciados por alguns fatores sob os quais destacam-se as multitarefas desenvolvidas pelos bibliotecários indexadores, a mecanicidade na indexação e falta de conhecimento prévio do assunto.

Desta forma, compreendendo que a leitura técnica para análise documentária é executada pela maioria dos participantes da pesquisa, coube verificar quais partes do documento eles costumam utilizar para a retirada dos termos de indexação. As respostas relacionadas a este aspecto foram obtidas através da questão 12 e foram transcritas no quadro a seguir:

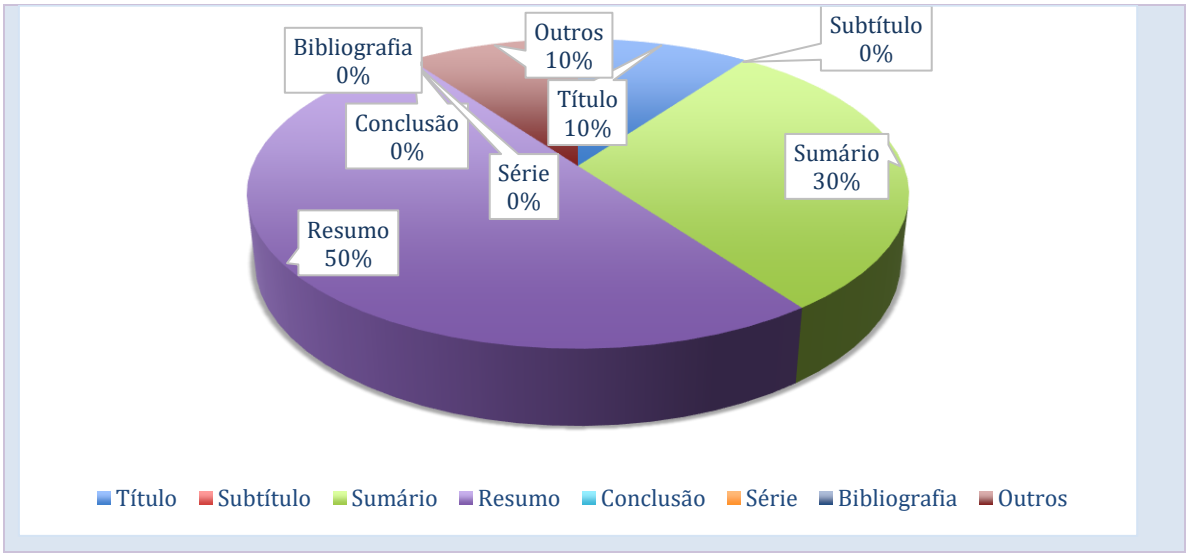
Quadro 18 - Partes do documento utilizadas pelos bibliotecários para análise documentária

Bibliotecário por ordem de resposta	Questão 12: **Quais as partes do documento você percorre para identificar os termos de indexação antes de traduzi-los a partir da linguagem controlada? Explique como ocorre esse procedimento.**
1	"Não se aplica"
2	"não se aplica"
3	"Não se aplica"
4	"Sumário e Resumo."
5	"Título, Resumo, Palavras-chave, sumário, apresentação (se houver), introdução."
6	"Normalmente atendo para a ficha catalográfica, resumo e introdução."
7	"Geralmente o resumo, sumário e apresentação"
8	"Resumo"

Fonte: elaborado pela autora, 2020.

A partir destas respostas e das recomendações feitas por Dias e Naves (2013), montou-se o gráfico 9, fazendo-se um comparativo entre o que recomenda a teoria e o que é adotado na prática pelos bibliotecários:

Gráfico 9 – Partes do texto recomendadas pela teoria e partes utilizadas pelos bibliotecários para análise documentária.



Fonte, elaborado pela autora, 2020.

Nota-se aí que nem todos os elementos citados pela teoria consultada estão presentes no dia a dia da prática bibliotecária, sendo que três destes elementos (50% do total) são

adotados pelos bibliotecários e outros três elementos (também 50% do total) não foram citados como preferidos para a leitura técnica. O termo preferido para a leitura e retirada dos termos de indexação é o resumo, sendo citado 5 vezes, seguido pelo sumário, que foi citado 3 vezes pelos bibliotecários. A apresentação, citada por Dias e Naves (2013) como um dos elementos principais para a leitura técnica, foi citada por apenas dois bibliotecários. A prevalência de resumo e sumário como elementos preferidos para a leitura técnica, conversa com o que afirma Redigolo (2014) quando a autora afirma que em algumas situações os indexadores selecionam termos sem selecionar conceitos. Resumo e sumário, embora contenham os elementos principais do assunto pesquisado, não abarcam toda a dimensão do seu real conteúdo.

Assim, diante das informações já obtidas sobre a análise de assunto, pode-se fazer o seguinte resumo:

a) os termos sugeridos pelos autores são levados em consideração pelos bibliotecários no momento da indexação;
b) a análise de assunto, por meio da leitura documentária, é adotada na rotina da indexação por parte dos bibliotecários;
c) resumo, sumário, apresentação e título são as partes mais citadas como sendo utilizadas pelos bibliotecários.

Percebe-se que o processo técnico de análise dos documentos a serem disponibilizados nos repositórios é dotado de etapas claras, a saber: conhecer o(s) assunto(s) contido(s) no texto e percorrer partes específicas do mesmo para, assim, ser feita a verificação de pertinência ao índice do catálogo.

Verificadas tais etapas, partiu-se para a observação do fenômeno da tradução conceitual, ou seja, o uso de termos autorizados por uma LD.

5.3.2 Tradução conceitual

Nesta categoria comportam-se as questões 14 e 15 que pretendem mensurar o que os bibliotecários julgaram ser mais importante no que se refere às linguagens documentárias, se a organização ou recuperabilidade. Além disso, essas questões pretendem captar quão importante eles consideram as LDs. Os quadros 19 e 20 apontam as respostas obtidas.

Quadro 19 - Importância das LDs

Bibliotecário por ordem de resposta	Questão 14: **Na sua opinião, qual a importância do uso de linguagens documentárias na tradução de termos para indexação de documentos digitais? Quais aspectos de sua utilização você considera mais importantes?**
1	"Não se aplica"
2	"não se aplica"
3	"é importante para fins de manter um padrão terminológico. O uso de remissivas"
4	"Como são utilizadas também a linguagem natural, acredito que por isso facilite muito mais na busca, assim o usuário encontra a informação seja por meio de linguagem natural ou termos específicos.?"
5	"Padronização das informações e recuperação da informação eficiente."
6	"A recuperação documental é o objetivo da utilização das LD'S. Por isso, os termos técnicos utilizados pelos autores e pesquisadores de uma determinada área devem ser levados em consideração."
7	"A garantia sob a perspectiva profissional da uniformidade e segurança no trabalho do catalogador e aprimorar a representação e recuperação temática de assuntos no catálogo online (BU ou RI)."
8	"Muito importante."

Fonte: elaborado pela autora, 2020.

De um modo geral pode-se considerar como um consenso que a utilização das LD'S possibilita um padrão, assim como facilita a recuperação de documentos com mais eficiência. Corroborando, deste modo, com o entendimento apresentado na seção 2.2.1 deste trabalho, sobre a leitura documentária caracterizar-se como atividade técnica, realizada com fins específicos, a fim de extrair conceitos que representem o conhecimento registrado. O objetivo, como afirma Fujita (2004), consiste na sintetização do conteúdo do documento. O quadro 20 demonstra os aspectos sinalizados pelos bibliotecários na questão 14, como é possível visualizar abaixo.

Quadro 20 - Aspectos importantes do uso das linguagens documentárias na tradução conceitual.

	ASPECTOS CITADOS PELOS BIBLIOTECÁRIOS
1	Padronização terminológica
2	Uso de remissivas
3	Em conformidade com a LN, facilita a busca pelo usuário
4	Recuperação mais eficiente
5	Uniformidade e consistência ao catálogo

Fonte: elaborado pela autora, 2020.

Já no Quadro 21 se apresentam as respostas sobre o aspecto organizacional e recuperabilidade.

Quadro 21 - Aspectos da utilização de linguagens documentárias: organização e recuperabilidade

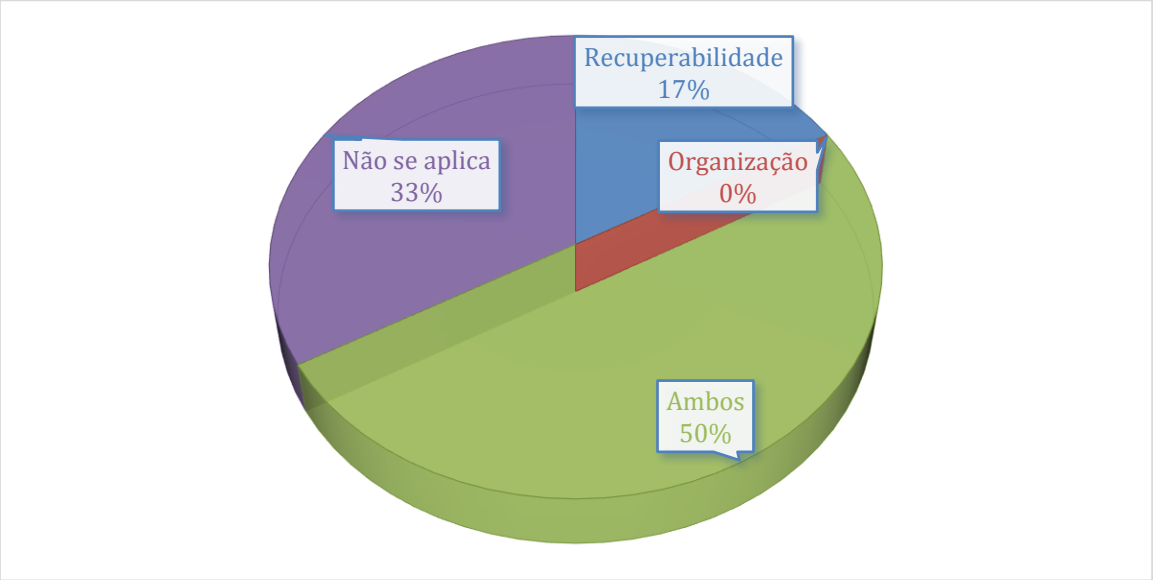
Bibliotecário por ordem de resposta	Questão 15: **Na etapa de seleção e tradução de conceitos, você considera mais importante o aspecto organizacional ou a recuperabilidade proporcionados pelas linguagens documentárias?**
1	"Não se aplica"
2	"não se aplica"
3	"depende do objetivo e da necessidade da instituição"
4	"Ambos"
5	"Recuperabilidade"
6	"Ambos são importantes e um deveria depender do outro. Porém, se eu tiver que escolher, prefiro a recuperabilidade."
7	"Quando se consegue tornar o catálogo uniforme e consistente, serve principalmente para amparar os indexadores no momento da representação temática dos documentos para a recuperação do documento."
8	"Recuperabilidade"

Fonte: elaborado pela autora, 2020.

A partir das repostas apresentadas no quadro 21, pode-se chegar na seguinte análise sobre as opiniões dos bibliotecários que atentam ao aspecto organizacional e à

recuperabilidade proporcionados pelas linguagens documentárias. Sendo assim, o gráfico 10 apresenta as opiniões dos bibliotecários sobre as questões levantadas.

Gráfico 10 - Aspectos da utilização das linguagens documentárias.



Fonte: elaborado pela autora, 2020.

Percebe-se que para metade dos bibliotecários respondentes os aspectos relacionados à organização e recuperabilidade são igualmente importantes. Enquanto que 17% acreditam que a recuperabilidade é o fator mais importante. Neste sentido, concorda-se com Guimarães (2009) quando afirma que a organização da informação é importante para as bibliotecas digitais e repositórios institucionais, pois é muito fácil que uma informação se perca em meio ao volume de informações veiculadas no meio digital.

A recuperação apresenta-se como fator mais importante, pois é ela quem traz o objeto que é a informação/documento. A organização contribui, ou pode contribuir, de forma satisfatória neste processo de recuperação, em síntese.

O Quadro 22 apresenta a questão 16, que aborda como é tratada a especificidade no repositório.

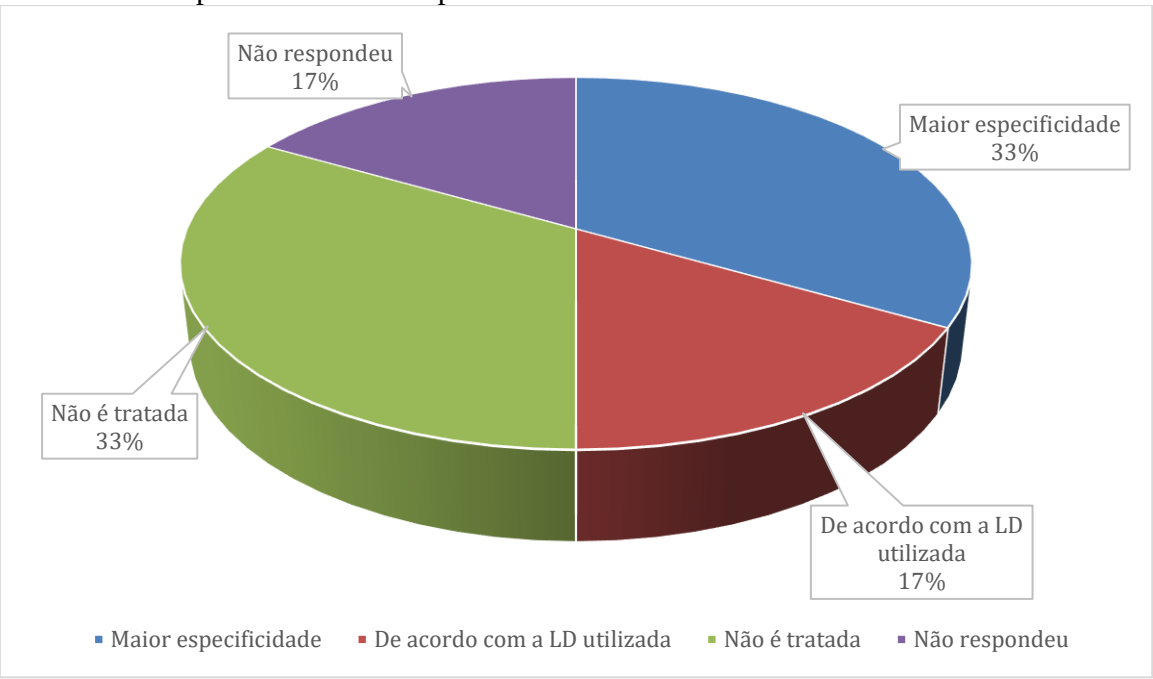
Quadro 22 - Especificidade nos repositórios

Bibliotecário por ordem de resposta	Questão 16: **Como é tratada a questão da especificidade no repositório?**
1	"Não é tratada"
2	"Não é tratada"
3	"Utilizar o termo mais específico que representa o conteúdo do documento"
4	Não respondeu
5	"Usamos os termos exaustivamente"
6	"Com base na CDU"
7	"Foi estabelecido parâmetros comuns - Termo Tópico - a serem adotados por todas as unidades de informação que formam a Rede do Sisbib tanto para BU e RI."
8	"Nosso repositório ainda é novo, ainda estamos aperfeiçoando."

Fonte: elaborado pela autora, 2020.

As respostas sofreram variações que equilibraram entre maior especificidade e sem especificidade. E entre as respostas ausentes e as de acordo com a LD também houve equilíbrio.

Gráfico 11 - Especificidade nos repositórios.



Fonte: elaborado pela autora, 2020.

O quadro a seguir suscita unanimidade nas respostas sobre a utilização de termos qualificadores.

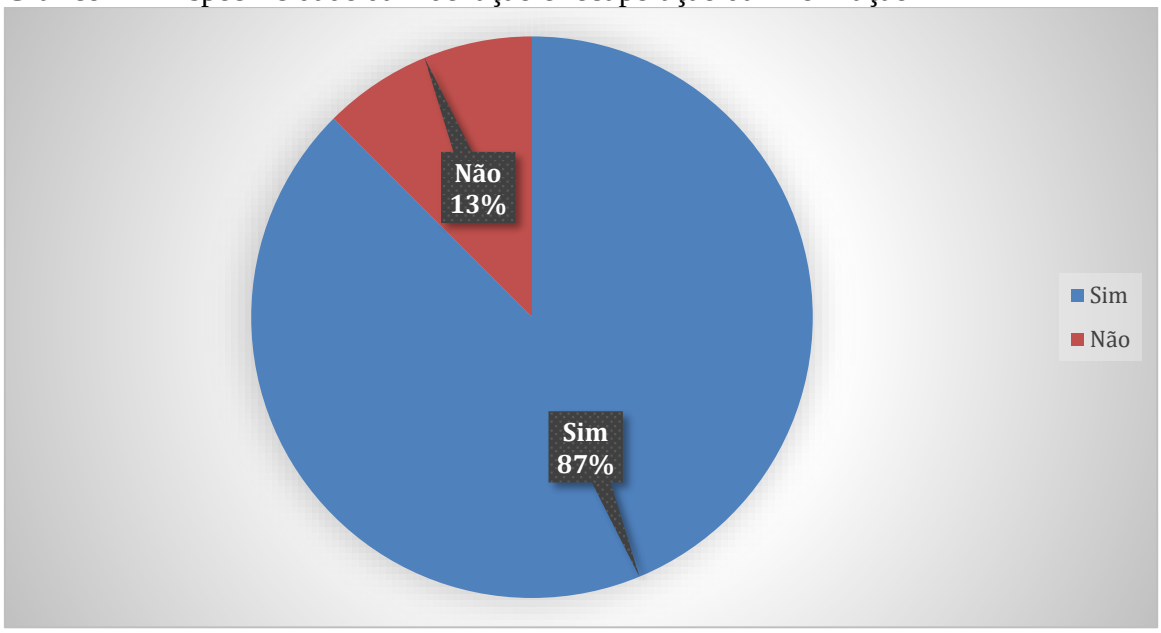
Quadro 23 - Uso de termos qualificadores e subdivisões

Bibliotecário por ordem de resposta	Questão 17: **Você concorda que o uso de termos qualificadores, também chamados de subdivisões, influenciam na recuperabilidade dos documentos, já que possibilitam uma indexação mais específica?**
1	"Sim"
2	"Sim"
3	"Sim"
4	"Sim"
5	"Sim"
6	"Sim"
7	"Sim"
8	"Não"

Fonte: elaborado pela autora, 2020.

Baseado no quadro acima, o Gráfico 12 apresenta a porcentagem das respostas obtidas como "sim" e "não":

Gráfico 12 - Especificidade da indexação e recuperação da informação



Fonte: elaborado pela autora, 2020.

“A expressão de um conteúdo em forma sintética e sem ambiguidades que efetivamente se pode falar em recuperação da informação” (NEET, 1981, p. 12). Os termos influenciam na recuperabilidade, bem como possibilitam a indexação mais específica. Essa utilização de termos mais qualificados proporciona consequência no trato e gerenciamento dos termos de recuperação dos documentos, pois possibilitam pesquisas bem-sucedidas, aumentando a eficiência e qualidade da indexação mais específica.

Sobre a exaustividade na representação da linguagem documentária, o quadro 24 apresenta as respostas da questão 18 que abordou essa temática.

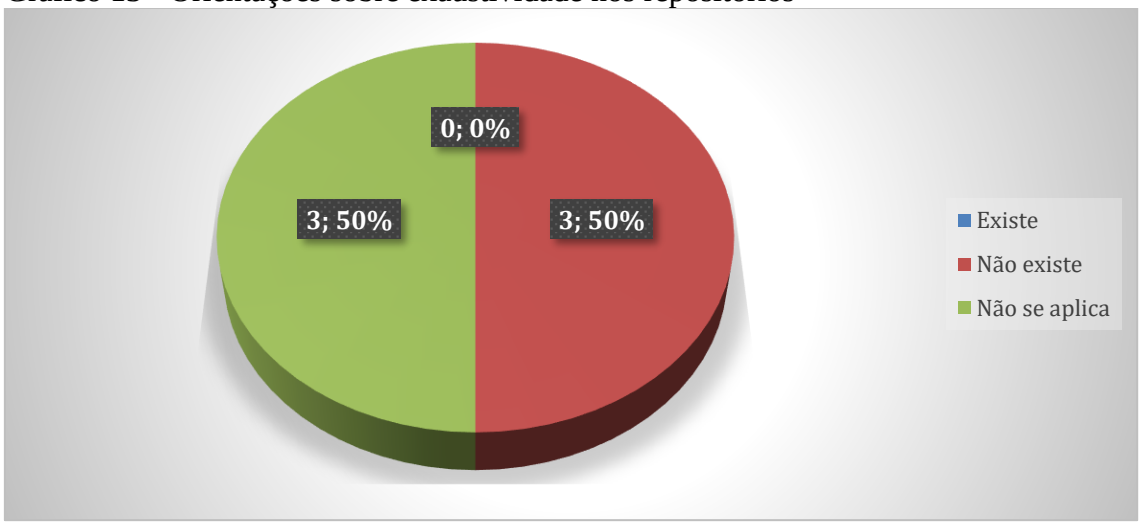
Quadro 24 - Orientações sobre a exaustividade nos repositórios.

Bibliotecário por ordem de resposta	Questão 18: **A exaustividade permite que todos os assuntos mais relevantes do documento sejam representados por meio de linguagens documentárias. Existem orientações quanto a essa questão no repositório no qual você trabalha? Quais?**
1	“Não se aplica”
2	“Não se aplica”
3	“Não se aplica”
4	“Não. Porém, como adotamos a CDU que possibilita trabalhar com especificidade exaustiva de assuntos, dependendo do indexadora a exaustividade ocorre.”
5	“Não há”
6	“Não. Porém, como adotamos a CDU que possibilita trabalhar com especificidade exaustiva de assuntos, dependendo do indexadora a exaustividade ocorre.”
7	“Não existe orientação neste sentido”.
8	“Não tem orientação”.

Fonte: elaborado pela autora, 2020.

A partir destas respostas, montou-se o Gráfico 13, a fim de melhor observar os aspectos pretendidos.

Gráfico 13 - Orientações sobre exaustividade nos repositórios



Fonte: elaborado pela autora, 2020.

Quando perguntados sobre a exaustividade e se ela permite que os assuntos mais relevantes do documento sejam representados através de linguagens documentárias, bem como se há orientações quanto a essa questão no repositório no qual trabalham, as respostas apontam para uma negativa de exaustividade.

Desta forma, o Quadro 25 traz as respostas recebidas diante da questão 19, que aborda o uso de remissivas nos repositórios.

Quadro 25 - Uso de remissivas nos repositórios.

Bibliotecário por ordem de resposta	Questão 19: **Uma pesquisa realizada por meio de um termo não autorizado em catálogos online de bibliotecas retorna documentos indexados sob determinado assunto. Isto se deve ao fato de catálogos online usarem a estratégia de remissivas em seu sistema, o que não é possível ainda nos repositórios, podendo, assim, causar dispersão documental ou ocultação de documentos na pesquisa. Quais os procedimentos adotados por você para lidar com esta questão?**
1	"Se repetem os termos relacionados."
2	"Se repetem os termos relacionados."
3	"Se for considerado o termo remissivo significativo, o acrescenta também."
4	"As palavras que poderíamos utilizar como remissivas normalmente são aquelas que os autores incluem nas suas

	palavras-chave, mas se houvesse um campo para as remissivas, seriam bem melhor realmente.”
5	“Usamos os termos qualificadores para especificar melhor.”
6	“Reeditar os termos indexados até a promoção de uma melhor busca.”
7	“Devido à metodologia de recuperação adotada e à elaboração de política para melhorar o sistema de recuperação, então os procedimentos estão em fase de elaboração.”
8	“Ainda não é possível descrever os procedimentos.”

Fonte: elaborado pela autora, 2020.

Baseando-se nas respostas dos bibliotecários entrevistos, pode-se entender que a utilização das remissivas coincidem com o objetivo da indexação, que é ajudar na recuperação da informação/documento. O entendimento sobre o qual se tem consenso é que pode-se acrescentar as remissivas onde já existe procedimento para tal, propiciando melhor tomada de decisão visando a busca mais precisa.

Quadro 26 - Estratégias utilizadas pelos bibliotecários em relação às remissivas nos repositórios.

	ESTRATÉGIAS RETIRADAS DAS RESPOSTAS DOS AUTORES
1	Repetição dos termos relacionados
2	Considera-se a importância da remissiva
3	Utilização das palavras do autor (LN) como remissivas
4	Uso de termos qualificadores
5	Reedição dos termos buscando atingir a maior qualidade na recuperação

Fonte: elaborado pela autora, 2020.

Pode-se perceber que, apesar do software utilizado pelos repositórios não executar por si só as remissivas, trazendo à tona no resultado da pesquisa documentos indexados de forma diferente ou relacionada, este aspecto é considerado pelos bibliotecários como parte importante no que tange à recuperabilidade da informação. Estratégias como o uso de termos qualificadores e termos relacionados para que um documento possa ser mais bem recupera do, demonstram bem a preocupação por parte dos bibliotecários quanto ao aspecto das remissivas.

Assim, esta questão supracitada traz à tona uma questão seguinte: o uso de termos traduzidos para outros idiomas como estratégia para aumentar a recuperabilidade da informação. Com o intuito de verificar este aspecto na atividade dos bibliotecários respondentes, destinou-se a pergunta de número 20, da qual as respostas estão transcritas a seguir no quadro 27.

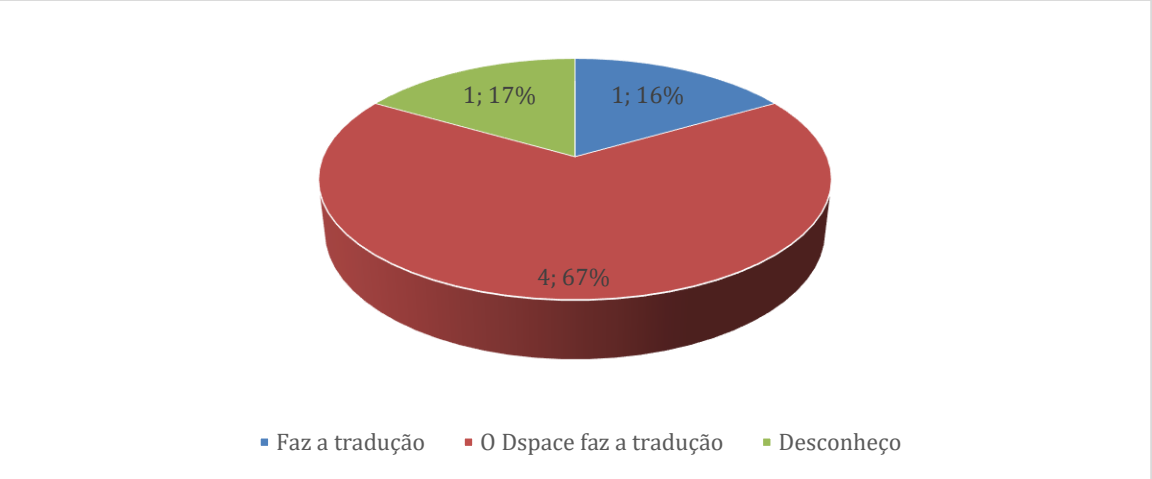
Quadro 27 - Tradução dos termos para outros idiomas.

Bibliotecário por ordem de resposta	Questão 20: **Os repositórios são ferramentas interoperáveis que formam uma rede de informação em âmbito mundial. Assim, no repositório que você trabalha, existe algum direcionamento para a tradução dos termos autorizados para outro idioma ou a tradução fica a cargo do próprio software (DSpace)?**
1	"Conforme consta no trabalho, o Dspace somente faz a tradução do interface"
2	"Conforme consta no trabalho, o Dspace somente faz a tradução do interface"
3	"O termo em diferentes idiomas é indexado quando o próprio documento apresenta esses termos. O Dspace traduz apenas a interface."
4	"Desconheço"
5	"Usamos os termos de indexação também no idioma inglês ou espanhol (de acordo com o documento)"
6	"A tradução fica a cargo do DSpace"
7	"Não sei informar"
8	"Desconheço"

Fonte: elaborado pela autora, 2020.

Desta forma, as respostas transcritas no quadro 27 possibilitaram a elaboração do gráfico 14, como ilustrado a seguir:

Gráfico 14 - Tradução dos termos para outros idiomas.



Fonte: elaborado pela autora, 2020.

Como representado no Gráfico 14, a maioria dos bibliotecários utilizam o Dspace para tradução. Sem maiores problemas, alguns o utilizam apenas para tradução da interface, mas a maioria deixa a cargo do software a tradução dos termos para outros idiomas. 17% desconhecem esta possibilidade ou se há a tradução para língua estrangeira, e 16% fazem a tradução de forma autônoma.

A questão 21, representada no quadro 28, procurou identificar se já houve resistência de autores quanto á tradução para a linguagem documentária.

Quadro 28 - Resistência dos autores em relação à tradução conceitual.

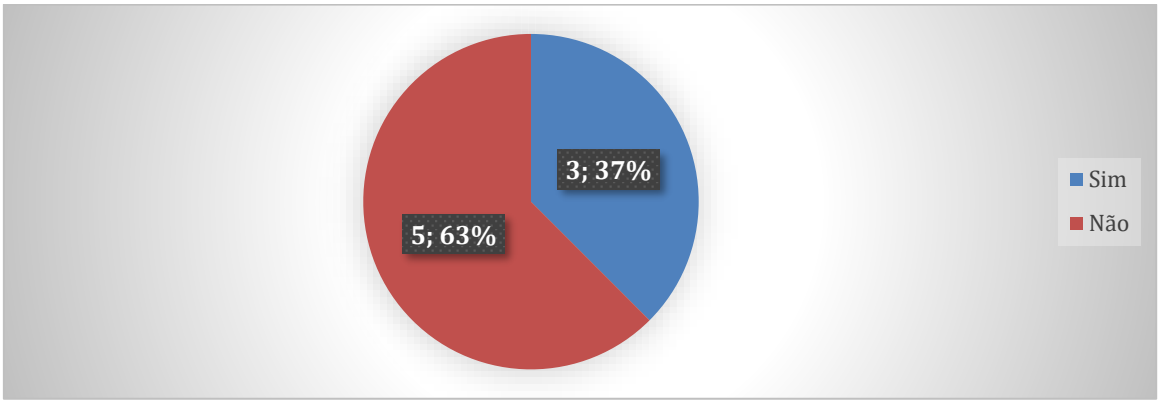
Bibliotecário por ordem de resposta	Questão 21: **Você já enfrentou resistência por parte dos autores quanto à tradução para uma linguagem documentária dos termos por eles utilizados? Se SIM, como solucionou este problema?**
1	"Não"
2	"Não"
3	"Não tive essa experiência"
4	"Sim. Aconteceu um caso. Umas das palavras que o usuário inseriu era muito "solta". Substituímos o termo por um mais específico. Após reclamação do autor, optamos por deixar ambos os termos."
5	"Sim. Esclarecemos que o trabalho terá uma melhor recuperação se for usado termos mais condizentes com a pesquisa do autor."

6	“Sim. Costumo explicar que o uso de LD"S promovem uma melhor recuperação da informação.”
7	“Ainda não”
8	“Não”

Fonte: elaborado pela autora, 2020.

As respostas obtidas através da questão 21 foram traduzidas em forma de gráfico (14), a fim de melhor visualizar os aspectos observados.

Gráfico 15 - Resistência dos autores em relação à tradução conceitual.



Fonte: elaborado pela autora, 2020.

Mais da metade dos respondentes (63%) não teve problemas relacionados com os autores, enquanto o restante (37%) já apresentou pelo menos uma situação de resistência. As medidas tomadas partem do entendimento e bom senso, às vezes, acrescentando o termo do autor, e outras vezes explicando a razão da utilização das linguagens documentárias.

Por fim, foram destinadas duas perguntas para que se pudesse observar questões referentes à região específica onde os estados do Norte do Brasil estão situados, a Amazônia. A questão 22, representada no quadro 29, aborda a temática da indexação de documentos científicos na Região Amazônica.

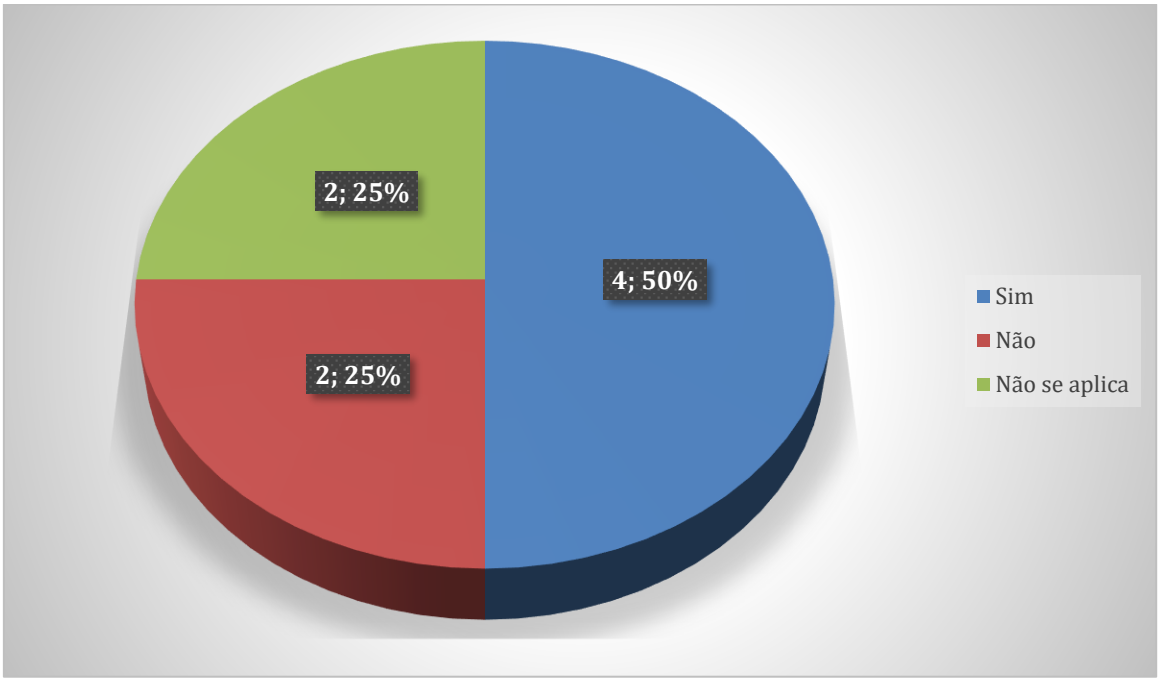
Quadro 29 - Dificuldades enfrentadas na indexação de documentos científicos na Amazônia.

Bibliotecário por ordem de resposta	Questão 22: **A construção e atualização das linguagens documentárias é baseada na garantia literária (termos mais citados nos textos científicos de determinada área do conhecimento) e isto pode implicar na ausência de termos indexadores para documentos produzidos em contextos específicos, como os culturais e geográficos. Dessa forma, você já sentiu dificuldades para indexar pesquisas relativas à região amazônica, da qual a Região Norte do Brasil faz parte, sendo que ainda são poucas e limitadas as linguagens documentárias voltadas especificamente para este local? Se SIM, como solucionou o problema?**
1	"Não se aplica"
2	"Não se aplica"
3	"Com o uso de linguagem natural"
4	"Isso acontece com muita frequência. Procuramos substituir o termo por um mais próximo possível. A subdivisão geográfica ajuda muito nesses casos."
5	"No nosso caso, que usamos a linguagem natural, não há dificuldades pois adequamos os assuntos para nossa realidade."
6	"Sim. Porém, Quem mora na região norte costuma adotar linguem natural sobre as questões ou fenômenos locais. Diante disso, o usuário "primário" dessa informação pode ser consultado. Antes da padronização de uma linguagem, podemos consider (*sic*) as expressões populares."
7	"Sim, mas estamos discutindo no âmbito da elabora (*sic*) da política de indexação institucional"
8	"Ainda não indexamos pesquisas relativas à Região Norte"

Fonte: elaborado pela autora, 2020.

As respostas transcritas no Quadro 28 possibilitaram a elaboração do gráfico 16, como ilustrado abaixo:

Gráfico 16 - Dificuldades enfrentadas na indexação de pesquisas referentes à Região Norte/Amazônia.



Fonte: elaborado pela autora, 2020.

Quando perguntados sobre as dificuldades enfrentadas para a indexação de documentos científicos na Amazônia, as respostas se dividiram de forma heterogênea, mas apontando uma maioria com dificuldades para esta atividade, pelo menos 50% dos respondentes afirmaram ter dificuldades para indexação destas pesquisas. A busca por soluções levou alguns bibliotecários a utilizarem a linguagem natural como recurso para sanar a ausência de termos locais nas LDs gerais e mais utilizadas.

Desta forma, o quadro 30, elaborado a partir das respostas dadas à questão 23, traz uma breve abordagem sobre a visão que os bibliotecários têm sobre a atualização das LDs no tocante aos termos científicos originários da Amazônia.

Quadro 30 - Atualização das LDs em relação aos termos científicos na Amazônia.

Bibliotecário por ordem de resposta	Questão 23: **Levando em consideração o rápido desenvolvimento do conhecimento científico e a necessidade de sua disseminação e recuperação, especialmente online, você considera que a atualização das linguagens documentárias acompanham esse desenvolvimento para indexação dos documentos? Justifique.**

1	"Não se aplica"
2	"Não se aplica"
3	"Depende da rapidez em que ocorre a atualização da linguagem documentária"
4	Não respondeu
5	"Não. Pois não há em constante atualização para as diferentes áreas do conhecimento."
6	"Em se tratando da CDU que utilizo datada de 2007, ferramenta que adoto para a indexação de documentos. Não acompanha o rápido desenvolvimento do conhecimento científico."
7	"Sim. Estão em constante mudança"
8	"?"

Fonte: elaborado pela autora, 2020.

A negativa para esta questão foi unânime. A maioria dos bibliotecários entrevistados considerou que as atualizações não acompanham de forma rápida o desenvolvimento científico, incorporando novos termos surgidos em pesquisas em regiões cheias de particularidades como a Amazônia. Solução para esta questão seria a adoção da garantia cultural, citada por Barité (2007), onde autor aponta que esta deve ser uma extensão da garantia literária para inserção de termos em linguagens documentárias, pois ela garante a inclusão de termos que respeitem características culturais ou geográficas.

Após a análise das perguntas de forma individual, partiu-se para uma análise global do estudo, abordando os principais pontos observados.

5.4 Discussão geral dos resultados

A questão guia desta pesquisa foi a verificação do uso de LDs para a indexação de documentos digitais em repositórios das universidades federais na Região Norte do Brasil. Abordando diversas questões referentes a esta investigação, pode-se observar importantes aspectos relacionados ao uso de LDs em tais repositórios. O Quadro 31 traz a relação das categorias de análise com os resultados obtidos, como se pode ver abaixo:

Quadro 31 - Relação das categorias de análise e dos resultados obtidos.

N. DA SESSÃO MACRO OBJETIVA	OBJETIVO	CATEGORIAS DE ANÁLISE	RESULTADOS OBTIDOS
1	Ambientação do repositório e das atividades dos bibliotecários	Identificação do repositório e das atividades do indexador	✔ as universidades federais da Região Norte estão, em sua maioria, inseridas no contexto de expansão do livre acesso às pesquisas científicas no Brasil; ✔ os Bibliotecários residentes no norte do país e que trabalham com indexação de documentos em repositórios não têm esta atividade como principal atribuição, desenvolvendo-a em meio a outras atividades; ✔ a maioria dos bibliotecários também não se dedica à indexação de documentos de uma área específica do conhecimento.
2	Analisar a adoção ou não adoção de política de indexação e uso de linguagens documentárias na indexação de documentos digitais	Política de indexação e Linguagens documentárias Uso de Linguagen Natural e Linguagem Documentária (combinada) Indexação de documentos digitais	✔ as política de indexação, que são instrumentos fundamentais para a determinação do uso das LDs, não estão presentes na grande maioria dos repositórios; ✔ na ausência de uma política de indexação, os bibliotecários tendem a optar pela utilização de termos livres e linguagem natural; ✔ não existindo uma política que clarifique os cuidados a serem tomados com a indexação de documentos digitais, estes são submetidos aos mesmos procedimentos adotados com o acervo físico da biblioteca, deixando de considerar as

			particularidades da recuperação digital;
3	Analisar a tradução conceitual de termos para linguagens documentárias	Análise de assunto Tradução conceitual Especifçidade na indexação Exaustividade na indexação Indexação e reuperabilidade Adequação das linguagens documentárias utilizadas e a indexação de documentos científicos na Região norte	✔ na análise de assunto, os termos utilizados pelos autores são considerados pelos bibliotecários para a indexação do documento; ✔ a análise documentária é desenvolvida a partir da leitura técnica de partes específicas do texto, como título e resumo; ✔ apesar da maioria dos bibliotecários adotar a LN, os mesmos entendem a importância das LDs como instrumentos de padronização terminológica; ✔ a recuperabilidade é considerada o ponto principal da tradução conceitual; ✔ especificidade e exaustividade são vistos como elementos importantes na recuperabilidade informacional; ✔ a LN, utilizada pela maioria dos respondentes, também é incorporada como alternativa ao uso de remissivas; ✔ o uso de LDs, que permite um controle terminológico na indexação e melhoria na recuperação da informação, é desconhecido por parte dos autores/pesquisadores que, em algumas situações, costumam questionar a “troca” de suas palavras-chaves; ✔ termos específicos originários da região amazônica, apesar de contemplados em grande parte pelas LDs gerais

			existentes, ainda figuram como um entrave na tradução conceitual para uma linguagem especializada. Como alternativa a esta questão, a LN é utilizada.

Fonte: elaborado pela autora, 2020.

Os pontos observados na macrossessão 1 mostram que a maioria das universidades federais da Região Norte do Brasil já possuem repositórios e que a indexação de documentos nestes ambientes é uma dentre tantas atividades desenvolvidas pelos bibliotecários. Isto levanta determinada preocupação em relação à maneira que a indexação vem sendo desenvolvida, podendo tornar-se apenas mais uma dentre as tantas atribuições dos bibliotecários que, muitas vezes sobrecarregados pelo trabalho, podem mecanizar a atividade de indexação, tornando-a algo "automático".

Já os aspectos observados através da macrossessão 2, demonstram que a adoção de LDs para indexação de documentos em repositórios digitais ainda não é uma unanimidade formalizada entre as instituições mantenedoras dos repositórios. Isto é reflexo da ausência de diretrizes que conduzam esta atividade, como é o caso de uma política de indexação. Outro ponto verificado na macrossessão 2 foi que os bibliotecários costumam adotar uma média de 1 a 6 termos para representação dos documentos. Este aspecto, caso se torne automático, apenas para preencher determinado número de termos, pode refletir tanto na especificidade quanto na exaustividade, já que se observou a ausência de uma política que parametrize os procedimentos a serem adotados relativos a estas questões.

Por fim, a macrossessão 3 explicita o uso da LN como principal instrumento de indexação dos documentos nos repositórios federais universitários da Região Norte, apesar do entendimento que os responsáveis por estes repositórios têm sobre a importância do uso de LDs para o controle terminológico. Porém, o uso da LN não suscita apenas aspectos negativos como a dispersão terminológica e, consequentemente, a documentária, mas também, sendo utilizada com o devido cuidado, é capaz de solucionar problemas como a urgente necessidade de utilização de termos já estabelecidos entre os cientistas para a indexação de documentos.

6 CONSIDERAÇÕES FINAIS

O volume cada vez maior, os novos formatos e suportes, assim como a necessidade de se encontrar uma informação pertinente de maneira rápida, são fatores que influenciaram, dentro da área da Organização da informação, o desenvolvimento de novas ferramentas, assim como a atualização dos já existentes (listas de cabeçalhos de assuntos, tesauros, etc) que possibilitassem maior controle na organização dos acervos das bibliotecas.

A organização do conhecimento, ao possibilitar a reunião de conceitos dentro de uma área científica, é a base para a construção de sistemas complexos, dotados de propriedades semânticas e lógicas. Estes sistemas são chamados de SOCS e funcionam como ferramentas que auxiliam a organização da informação, esteja ela em acervo físico ou digital. O digital como meio de armazenamento e transmissão da informação, causou verdadeira revolução na forma de tratar a organização de um acervo, pois o foco transferiu-se do mero ordenamento para a recuperação da informação contida nos documentos.

Assim, as linguagens documentárias também se modificaram para atender a uma nova demanda informacional, a da recuperação. Demanda esta que necessitava também de novos meios de transferência e acesso, permitindo que pessoas de diferentes localidades pudessem usar a informação disponibilizada em diversos lugares, próximos ou distantes geograficamente. Os repositórios digitais surgiram desta necessidade, assim como também de abrigar um grande volume de informações, coisa que fisicamente não seria possível sem a disponibilização de grandes espaços, já que a pesquisa científica está em constante transformação.

As universidades federais públicas são responsáveis por inúmeras pesquisas científicas desenvolvidas nas diversas áreas do conhecimento, pois abrigam cursos das ciências humanas, exatas, naturais e sociais. Neste sentido, são guardiãs e disseminadoras de informação de qualidade, desenvolvidas por seus docentes, discentes, funcionários e colaboradores.

O acesso livre aos resultados destas pesquisas permite que a comunidade extra universitária possa consumir os produtos gerados pelo investimento público em pesquisas científicas, base de desenvolvimento dos grandes países. Os repositórios digitais universitários são ferramentas que nasceram e se espalharam como alternativa ao controle financeiro das grandes editoras sobre as pesquisas, proporcionando à ampla comunidade o acesso a elas.

Porém, não adianta apenas povoar o espaço dos repositórios digitais universitários com os documentos ali disponibilizados, é necessário que estas informações estejam tratadas de maneira a serem facilmente encontradas por quem delas necessita. A indexação, quando bem executada, é um procedimento que permite que os documentos sejam habilmente recuperados pelo sistema, reunindo-os de forma satisfatória no momento da pesquisa.

A utilização das linguagens documentárias, através de termos autorizados, é ponto de convergência da representação da informação, pois é ela quem evita que termos conceituados de formas diferentes em cada área sejam representados da mesma maneira, assim como evita que diversos conceitos sobre o mesmo termo sejam representados de maneira diferente. Isto facilita não só a recuperação da informação dentro do próprio sistema, mas também influencia na visibilidade do catálogo da instituição dentro de sistemas mais amplos, como a BDTD.

Sendo a região amazônica um rico campo de trabalho para os pesquisadores, se faz necessário que as universidades aqui residentes assumam o papel não só de geradoras, mas também de divulgadoras de todo o conhecimento por elas gerado na região. Para isso, há que se respeitar as particularidades de termos científicos aqui desenvolvidos, muito embora a atualização das linguagens por meio da garantia literária não alcance o rápido desenvolvimento das pesquisas. Em consonância com esta afirmativa, pode-se concluir que a utilização da linguagem natural, por meio de termos ainda não inseridos nas linguagens documentárias, possam sanar momentaneamente essa lacuna até que os termos sejam inseridos como autorizados por uma LD.

Ao ter constatado a prevalência da linguagem natural, percebeu-se que a ausência da política de indexação na maioria dos repositórios investigados influenciou o resultado. Assim, este trabalho se propõe a servir como uma semente para novas investigações sobre a indexação de documentos em espaços digitais como os repositórios e também sobre a adoção de política que guie este procedimento, em especial na região onde foi realizada a pesquisa.

REFERÊNCIAS

AMERICAN NACIONAL STANDARDS ORGANIZATION – ANSI. ANSI/NISO Z 39.19: 2005: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda: NISO Press, 2005. 176p. Disponível em: https://groups.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf. Acesso em: 17 jul. 2019.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR 12676**: métodos para análise de documentos: determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro: ABNT, 1992.

BARITÉ, Mário. La garantia literária: vigência y proyección teórico-metodológica. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8, Salvador. Anais... Salvador: ANCIB, 2007. p. 1-10. Disponível em: http://enancib.ibict.br/index.php/enancib/viiienancib/search/authors/view?firstName=Mario&middleName=&lastName=Barité&affiliation=&country=. Acesso em: 06 jan. 2020.

BENTES PINTO, Virgínia. Indexação documentária: uma forma de representação do conhecimento registrado. **Perspec. Info.**, v. 6, n. 2, p. 223-234, jul./dez. 2001.

BOCCATO, Vera Regina Casari. A linguagem documentária vista pelo conteúdo, forma e uso na perspectiva de catalogadores e usuários. In: FUJITA, M. S. L., et al. (org.). **A indexação de livros:** a percepção de catalogadores e usuários de bibliotecas universitárias. Um estudo de observação do contexto sociocognitivo com protocolos verbais [online]. São Paulo: Editora UNESP; São Paulo: Cultura Acadêmica, 2009. 149 p.

BRASCHER, M., CAFÉ, L. Organização da informação ou organização do conhecimento? In: **Enancib - Encontro Nacional de Pesquisa em Ciência da Informação**, 2008, São Paulo. Diversidade Cultural e Políticas de Informação, 2008.

BRASIL. Ministério da Ciência e Tecnologia. Termo de Compromisso de Gestão – TCG. Brasília: MCT, 2007. Disponível em: https://www.mctic.gov.br/mctic/export/sites/institucional/contratos_gestao_organizacoes_sociais/arquivos/TCG/IBICT/Termo-de-Compromisso-de-Gestao-2007.pdf. Acesso em: 19 jul. 2019.

BRASIL. Ministério da Educação. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Portaria n. 013, de 15 de fevereiro de 2006**. Institui a divulgação digital das teses e dissertações produzidas pelos programas de doutorado e mestrado reconhecidos. Brasília: CAPES, 2006.

BUDAPESTE OPEN ACCESS INITIATIVE. [S. l.: s. n.], 2002. Disponível em: https://www.budapestopenaccessinitiative.org/read. Acesso em: 18 jul. 2019.

BUDAPESTE OPEN ACCESS INITIATIVE. [S. l.: s. n.], 2012. Disponível em: https://www.budapestopenaccessinitiative.org/boai-10-translations/portuguese-brazilian-translation. Acesso em: 18 jul. 2019.

BUDAPESTE OPEN ACCESS INITIATIVE. Disponível em: https://www.budapestopenaccessinitiative.org/read. Acesso em: 17 jul. 2019.

BUSH, V. As we may think. **Atlantic Monthly**, v. 176, n. 1, p.101-108, 1945. Disponível em: http://web.mit.edu/STS.035/www/PDFs/think.pdf. Acesso em: 15 jan. 2019.

CARLAN, Eliana. **Sistemas de organização do conhecimento**: uma reflexão no contexto da Ciência da Informação. Orientadora: Marisa Brascher Basílio Medeiros. 2010. 195 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2010. Disponível em: http://eprints.rclis.org/14519/1/Carlan-Eliana-Dissertacao.pdf. Acesso em: 29 jul. 2019.

CARLAN, Eliana; MEDEIROS, Marisa Brascher Basílio. Sistemas de Organização do Conhecimento na visão da Ciência da Informação. **Revista Ibero-Americana de Ciência da Informação**, [S.l.], v. 4, n. 2, fev. 2012. Disponível em: http://periodicos.unb.br/index.php/RICI/article/view/6209. Acesso em: 19 maio 2019.

CAVALCANTI, C. R. O. Indexação. **Estudos Avançados em Biblioteconomia e Ciência da Informação**, v. 1, n. 1, 1982. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/40406>. Acesso em: 07 jan. 2020.

CESARINO, Maria Augusta da Nóbrega; PINTO, Maria Cristina Mello Ferreira. Cabeçalho de assunto como linguagem de indexação. **Revista da Escola de Biblioteconomia da UFMG**, v. 7, n. 2, 1978, p. 268-288. Disponível em: http://www.brapci.inf.br/index.php/article/view/0000001991/41d251f63e7744e1fee36986a3780073/. Acesso em: 22 fev. 2019.

CHAUMIER, Jacques. Indexação: conceito, etapas e instrumentos. **Revista Brasileira de Biblioteconomia e Documentação**, v. 21, n. 1/2, p. 63-79, jan./jun. 1978. Disponível em: http://www.brapci.inf.br/index.php/article/view/0000011407/247399c870111947e2009836ea74fb3e. Acesso em: 14 de mar. 2019.

CONGRESSO MUNDIAL DE INFORMAÇÃO EM SAÚDE E BIBLIOTECAS, 9.; REUNIÃO DE COORDENAÇÃO REGIONAL DA BVS, 4., 2005, Salvador. **Declaração de Salvador sobre o acesso aberto**. Salvador: ICML/CRICS, 2005. Disponível em: http://www.icml9.org/channel.php?lang=pt&channel=86&content=428. Acesso em: 29 jul. 2019.

CROW, Raym. The case for institutional repositories: a SPARC position paper. **ARL Bimonthly Report**, n. 223, aug. 2002.

CUNHA, Isabel M. R. Ferin. Análise documentária. In: SMIT, J. W. (Coord.). **Análise documentária:** a análise da síntese. Brasília: IBICT, 1987. Disponível em: http://livroaberto.ibict.br/bitstream/1/1011/1/An%C3%A1lise%20document%C3%A1ria.pdf. Acesso em: 27 ago. 2019.

DAHLBERG, Ingetraut. Teoria do conceito. Ciência da Informação, [S. l.], v. 7, n. 2, dez. 1978. Disponível em: http://revista.ibict.br/ciinf/article/viewq/115. Acesso em: 19 maio 2019.

DECLARAÇÃO de apoio ao acesso aberto à literatura científica: carta de São Paulo. São Paulo: [s. n.], 2005. Disponível em: http://www.forum-global.de/acessoaberto/carta_de_sao_paulo_acesso_aberto.htm. Acesso em 29 jul. 2019.

DIAS, Eduardo Wense; NAVES, Madalena Martins Lopes. **Análise de assunto:** teoria e prática. 2. ed. rev. Brasília: Briquet de Lemos, 2013. 115 p.

DUBLIN CORE METADATA INITIATIVE. **Conjunto de metadados Dublin Core, versão 1.1**: descrição de referência. [S. l.]: DCMI, 2012. Disponível em: http://www.dublincore.org/documents/dces/. Acesso em: 19 maio 2019.

FAUSTO, S. Evolução do Acesso Aberto – breve histórico [online]. **SciELO em Perspectiva**, 2013. Disponível em: https://blog.scielo.org/blog/2013/10/21/evolucao-do-acesso-aberto-breve-historico/. Acesso em: 29 jul. 2019.

FERREIRA, et al. Política para indexação do acervo do SISTEBIB/UFAM. UFAM: Manaus, 2017. Disponível em: https://biblioteca.ufam.edu.br/images/doc/politica_indexacao_revistaRM.pdf. Acesso em: 06 jan 2020.

FERREIRA, Sueli Mara S. P. et al. Repositório da Produção Científica CRUESP: mais do que um consórcio, um trabalho integrado da USP, UNESP e UNICAMP. *In*: CONFERÊNCIA LUSO-BRASILEIRA DE ACESSO ABERTO, 4., 2013, São Paulo. **Sessão de pôster** [...]. São Paulo: USP, 2013. Disponível em: https://bdpi.usp.br/bitstream/handle/BDPI/43861/Reposit%c3%b3rio%20da%20Produ%c3%a7%c3%a3o%20Cient%c3%adfica%20CRUESP%20mais%20do%20que%20um%20cons%c3%b3rcio,%20um%20trabalho%20integrado%20USP,%20UNESP%20e%20UNICAMP.pdf?sequence=1. Acesso em: 21 jul. 2019.

FREITAS, Marilia; LEITE, Fernando. Proposição de diretrizes para o depósito da produção científica em repositórios institucionais de acesso aberto baseada na visão de diferentes atores do sistema de comunicação científica. **Encontros Bibli: revista eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 23, n. 53, p. 96-109, set. 2018. Disponível em: https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p96. Acesso em: 27 ago. 2019.

FUJITA, M. N. S. L. Organização e representação do conhecimento no Brasil: análise de aspectos conceituais e da produção científica do ENANCIB no período de 2005 a 2007. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 1, n. 1, p. 1-32, 2008. Disponível em: http://www.brapci.inf.br/v/a/7781. Acesso em: 25 maio 2019.

FUJITA, Mariângela Spotti Lopes et al. Linguagens de indexação em bibliotecas universitárias: estudo analítico. **Informação & Informação**, v. 24, n. 1, p. 190-225, jan./abr. 2019. Disponível em: http://www.uel.br/revistas/uel/index.php/informacao/article/view/31771/pdf. Acesso em: 05 maio 2019.

FUJITA, Mariângela Spotti Lopes. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, v. 1, n. 1, p. 60-90, jul./dez. 2003.

FUJITA, Mariângela Spotti Lopes. Política de indexação para bibliotecas: funções e finalidades. *In*: FUJITA, Mariângela Spotti Lopes (Org.). Política de indexação para bibliotecas: elaboração, avaliação e implantação. Marília: Oficina Universitária; São Paulo, Cultura Acadêmica, 2016.

FUJITA, Mariângela Spotti Lopes; SANTOS, Luciana Beatriz Piovezan dos. Política de indexação em bibliotecas universitárias: estudo diagnóstico e analítico com pesquisa participante. **Transinformação**, Campinas , v. 28, n. 1, p. 59-76, abr. 2016 . Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862016000100059&lng=en&nrm=iso. Acesso em: 19 maio 2019.

FUNDAÇÃO DE CIÊNCIA, APLICAÇÕES E TECNOLOGIAS ESPACIAIS. **Edital de chamada FINEP/PCAL/XBDB n. 002/2009.** São José dos Campos: FUNCATE, 2009. 6 p. Disponíel em: https://kuramoto.wordpress.com/2009/04/20/edital-fineppcalxbdb-n%C2%BA-0012009-foi-revogado/. Acesso em: 19 jul. 2019.

GINSPARG, Paul. ArXiv at 20. **Nature**, n. 476, p. 145-147, aug. 2011. Disponível em: https://www.nature.com/articles/476145a. Acesso em: 11 jul. 2019.

GUIMARÃES, José Augusto Chaves. Abordagens teóricas do tratamento temático da informação (TTI): catalogação de assunto, indexação e análise documental. **IBERSID**, v. 3, p. 105-117, 2009. Disponível em: https://www.ibersid.eu/ojs/index.php/ibersid/article/view/3730. Acesso em: 18 jun. 2019.

HJORLAND, Birger (2015). Theories are knowledge organizing systems (KOS). Knowledge Organisation, v. 42, n. 2, p. 113-128. https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/conhecimento/. Acesso em: 19 maio 2019.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Manifesto brasileiro de acesso livre à informação científica**. [Brasília]: IBICT, 2005. Disponível em: http://livroaberto.ibict.br/Manifesto.pdf. Acesso em: 29 jul. 2019.

JAPIASSU, Hilton; MARCONDES, Danilo. Dicionário básico de filosofia. 3. ed. rev. e ampl. Rio de Janeiro: Zahar, 2001. Disponível em: http://raycydio.yolasite.com/resources/dicionario_de_filosofia_japiassu.pdf. Acesso em: 28 jul. 2019.

KURAMOTO, Helio. **Acesso Livre**: como tudo começou. [S. l.: s. n.]: 2012. Disponível em: https://kuramoto.wordpress.com/?s=como+tudo+come%C3%A7ou. Acesso em: 20 jan. 2019.

KURAMOTO, Helio. **Carta aberta à SBPC**. [S. l.: s. n.], 2006a. Disponível em: https://kuramoto.wordpress.com/2006/07/24/carta-aberta-a-sbpc/. Acesso em: 19 maio 2019.

KURAMOTO, Hélio. Implantação de repositórios institucionais em universidades e instituições de pesquisa no Brasil: do projeto ao processo. In: GOMES, Maria João; ROSA, Flávia (org.). **Repositórios institucionais:** democratizando o acesso ao conhecimento. Salvador: EDUFBA, 2010. 204 p.

KURAMOTO, Helio. Informação científica: proposta de um novo modelo para o Brasil. Ciência da Informação, v. 35, n. 2, p. 91-102, maio/ago. 2006b. Disponível em: http://revista.ibict.br/ciinf/article/view/1144/1305. Acesso em: 20 jan. 2019.

LE COADIC, Yves-François. **A Ciência da Informação**. Brasília, DF: Briquet de Lemos, 1996.

LEITE, F. et al. **Boas práticas para a construção de repositórios institucionais da produção científica**. Brasília: IBICT, 2012. Disponível em: http://livroaberto.ibict.br/bitstream/1/703/1/Boas%20pr%c3%a1ticas%20para%20a%20construu%c3%a7%c3%a3o%20de%20reposit%c3%b3rios%20institucionais%20da%20produ%c3%a7%c3%a3o%20cient%c3%adfica.pdf. Acesso em: 29 jul. 2019.

LEITE, Fernando César Lima. **Como gerenciar e ampliar a visibilidade da informação científica brasileira:** repositórios institucionais de acesso aberto. Brasília: IBICT, 2009. Disponível em: http://livroaberto.ibict.br/bitstream/1/775/4/Como%20gerenciar%20e%20ampliar%20a%20visibilidade%20da%20informa%C3%A7%C3%A3o%20cient%C3%ADfica%20brasileira.pdf. Acesso em: 29 jul. 2019.

LYNCH, Clifford A. Institutional Repositories: essential infrastructure for scholarship in the Digital Age. **ARL**, n. 226, p. 1-7, Feb. 2003. Disponível em: https://www.cni.org/wp-content/uploads/2003/02/arl-br-226-Lynch-IRs-2003.pdf. Acesso em: 2 jun. 2007

MAIMONE, G. D.; SILVEIRA, N. C.; TÁLAMO, M. F. T. G. A. M. Reflexões acerca das relações entre representação temática e descritiva. **Informação & Sociedade: Estudos**, v. 21, n. 1, p. 27-35, 2011. Disponível em: http://www.brapci.inf.br/v/a/10197. Acesso em: 19 maio 2019.

MARRA, Patrícia dos Santos Caldas. O papel das bibliotecas universitárias na comunicação científica: um estudo sobre os repositórios institucionais. **Encontros Bibli**, v. 17, n. esp. 2, p. 174-194, 2012.

MARRA, Patrícia dos Santos Caldas. Visibilidade dos repositórios institucionais brasileiros: análise de diretórios internacionais de acesso aberto. **Rev. Eletron. de Comum. Inf. Inov. Saúde**, v. 8, n. 3, p. 330-343, set. 2014. Disponível em; https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/672/1320. Acesso em: 20 jul. 2019.

MICHAELIS Dicionário Brasileiro da Língua Portuguesa. [S. l.: Melhoramentos], [2015]. Disponível em: http://michaelis.uol.com.br/busca?r=0&f=0&t=0&palavra=conhecimento. Acesso em: 28 ago. 2019.

MIRANDA, Marcos Luiz Cavalcanti de. **Organização e representação do conhecimento:** fundamentos teórico-metodológicos na busca e recuperação da informação em ambientes virtuais. 2005. 353 f. Tese (Doutorado em Ciência da Informação) - Universidade Federal do Rio de Janeiro / Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2005. Disponível em: http://ridi.ibict.br/handle/123456789/690. Acesso em: 29 abr. 2019.

MIRANDA, Marcos Luiz Cavalcanti de. **Organização e representação do conhecimento:** fundamentos teórico-metodológicos na busca e recuperação da informação em ambientes virtuais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 7., 2006, Marília. Anais... Marília: UNESP, 2006.

MOOERS, Calvin N. Zatocoding applied to mechanical organization of knowledge. **American Documentation**. v. 2, n. 1, p.20-32, 1951. Dispomível em: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.5090020107. Acesso em: 18 mar. 2019.

MUELLER, Suzana Pinheiro Machado. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, v. 35, n. 2, p. 27-38, maio/ago. 2006.

NOVELLINO, Maria Salet Ferreira. Instrumentos e metodologias de representação da informação. **Inf. Inf.**, Londrina, v. 1, n. 2, p. 37-45, jul./dez. 1996. Disponível em: http://www.brapci.inf.br/_repositorio/2010/05/pdf_0e3cc20139_0010458.pdf. Acesso em: 15 out. 2018.

NUNES, Cláudio Omar. Algumas considerações acerca da ausência de políticas de indexação em bibliotecas brasileiras. Biblios, Rio Grande, n. 16, p. 55-61, 2004. Disponível em: https://periodicos.furg.br/biblos/article/view/411/96. Acesso em: 07 ago. 2019.

PINHEIRO, Lena Vânia Ribeiro; PEREZ, Helena Dodd. Tesauro Brasileiro de Ciência da Informação. Rio de Janeiro; Brasília: IBICT, 2014. Disponível em: http://sitehistorico.ibict.br/publicacoes-e-institucionais/tesauro-brasileiro-de-ciencia-da-informacao-1/copy_of_TESAUROCOMPLETOFINALCOMCAPA24102014.pdf. Acesso em: 26 jun. 2019.

PINTO, Maria Cristina Mello Ferreira. Análise e representação de assuntos em sistemas de recuperação da informação: linguagens de indexação. **Rev. Esc. Biblioteconomia**, v. 14, n. 2, p. 169-186, set. 1985. Disponível em: http://www.brapci.inf.br/index.php/article/view/0000002635. Acesso em: 12 maio 2019.

REDIGOLO, Franciele Marques. **O processo de análise de assunto na catalogação de livros em bibliotecas universitárias: aplicação do protocolo verbal**. Orientadora: Mariângela Spotti Lopes Fujita. 2014. 262 f. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista Julio de Mesquita Filho, Faculdade de Filosofia e Ciências, Marília, 2014. Disponível em: https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/redigolo_fm_do_mar.pdf. Acesso em: 20 dez. 2019.

REDIGOLO, Franciele Marques. **O processo de análise de assunto na catalogação de documentos: a perspectiva sociocognitiva do catalogador em contexto de biblioteca universitária**. Orientadora: Mariângela Spotti Lopes Fujita. 2010. 179 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista Julio de Mesquita Filho, Faculdade de Filosofia e Ciências, Marília, 2010. Disponível em: http://www.athena.biblioteca.unesp.br/exlibris/bd/bma/33004110043P4/2010/redigolo_fm_me_mar.pdf. Acesso em: 20 dez. 2019.

ROSA, Flávia; GOMES, Maria João. Dos promotores aos utilizadores: estudos sobre o RepositóriUM. In: GOMES, Maria João; ROSA, Flávia (Orgs.). **Repositórios**

institucionais: democratizando o acesso ao conhecimento. Salvador: EDUFBA, 2010. p. 153-203. Disponível em: http://repositorium.sdum.uminho.pt/handle/1822/11417. Acesso em: 20 jan. 2019

RUBI, Milena Polsinelli. Política de indexação. In: LEIVA, Isidoro Gil; FUJITA, Mariângela S. L. (Org.). **Política de indexação**. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012.

RUBI, Milena Polsinelli. **Política de indexação para construção de catálogos coletivos em bibliotecas universitárias**. Orientadora: Mariângela Spotti Lopes Fujita. 2008.166 f. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2008. Disponível em: https://repositorio.unesp.br/bitstream/handle/11449/103388/rubi_mp_dr_mar.pdf?se. Acesso em: 21 jan. 2020.

SALES, Rodrigo de; CAFE, Lígia. Diferenças entre tesauros e ontologias. **Perspect. ciênc. inf.**, Belo Horizonte , v. 14, n. 1, p. 99-116, abr. 2009 . Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362009000100008&lng=en&nrm=iso. Acesso em: 15 jun. 2018.

SAUSSURE, Ferdinand de. **Curso de linguística geral**. 27. ed. São Paulo: Cultrix, 2006.

SHINTAKU, Milton; MEIRELLES, Rodrigo. Manual do DSpace: administração de repositórios. Salvador: EDUFBA, 2010. Disponível em: https://repositorio.ufba.br/ri/bitstream/ri/769/1/Manual%20do%20Dspace(2).pdf. Acesso em: 19 maio 2019.

SHINTAKU, Milton; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Bibliotecas e repositórios no processo de publicação digital. **Biblios:** revista do Instituto de Ciências Humanas e da Informação, v. 30, n. 1, p. 60-79, 2016. Disponível em: file:///D:/Downloads/5762-17938-1-PB%20(1).pdf. Acesso em: 20 jul. 2019.

SILVA, Maria dos Remédios da; FUJITA, Mariângela Spotti Lopes. A prática de indexação: análise da evolução de tendências teóricas e metodológicas. **Transinformação**, v. 16, n. 2, p. 133-161, maio/ago. 2004. Disponível em: https://repositorio.unesp.br/bitstream/handle/11449/114455/S0103-37862004000200003.pdf?sequence=1. Acesso em 22 fev. 2019.

SILVEIRA, Fernando Lang da. A teoria do conhecimento de Kant: o idealismo transcendental. **Caderno Brasileiro de Ensino de Física**, Florianópolis, p. 28-51, jan. 2002. Disponível em: https://periodicos.ufsc.br/index.php/fisica/article/view/10053/15383. Acesso em: 29 jul. 2019.

WEITZEL, S. R. O papel dos repositórios institucionais e temáticos na estrutura da produção científica. **Em Questão** [online], 2006a, n. 12, v. 1, p. 51-71. Disponível em: https://www.redalyc.org/pdf/4656/465645954004.pdf. Acesso em: 27 ago. 2019.

WEITZEL, S. R. **Os repositórios de e-prints como nova forma de organização da produção científica: o caso das áreas das Ciências da Comunicação no Brasil**. Orientadora: Sueli Mara dos Santos Pinto Ferreira. 360 f. 2006b. Tese (Doutorado em Cultura e Informação) – Escola de Comunicação e Artes, Universidade de São Paulo, São

Paulo, 2006. Disponível em: http://www.teses.usp.br/teses/disponiveis/27/27151/tde-14052009-133509/pt-br.php. Acesso em: 27 ago. 2019.

WEITZEL, S. R. Reflexões sobre os repositórios institucionais. *In*: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 29, 2006c, Brasília. **Anais** [...]. Brasília: UnB, 2006. Disponível em: http://eprints.rclis.org/8744/1/reflexoes_weitzel_endocom.pdf. Acesso em: 27 ago. 2019.

APÊNDICE

APÊNDICE A – Questionário: Utilização das Linguagens Documentárias (LDs) na tradução conceitual em repositórios da Região Norte do Brasil

Este questionário foi elaborado com o objetivo de compreender elementos de uso de linguagens documentárias alfabéticas nos repositórios institucionais universitários da Região Norte do Brasil. A sua contribuição é muito importante para traçarmos um panorama sobre a indexação de documentos em ambiente virtual, neste caso, dos repositórios.

1) IDENTIFICAÇÃO DO REPOSITÓRIO

2) Nome do repositório e instituição onde você desenvolve as atividades:

3) Ano em que o repositório entrou em atividade:

4) Você trabalha exclusivamente com indexação de documentos ou desenvolve outras atividades? Quais?

5) Você trabalha com a indexação de documentos de diversas áreas do conhecimento ou de uma área específica? Se sim, qual?

6) Caso seja em uma área específica, você costuma sentir dificuldade na tradução dos termos para uma linguagem documentária?

7) POLÍTICA DE INDEXAÇÃO E USO DAS LINGUAGENS DOCUMENTÁRIAS

8) Existe, em sua instituição, uma política de indexação que determine o uso de linguagens documentárias?

9) Caso a resposta da pergunta 6 seja SIM, sobre o uso de linguagens documentárias, quantas e quais linguagens ela recomenda? Caso seja NÃO, quantas e quais você costuma utilizar e por quê?

10) Ainda sobre política de indexação, caso exista em sua instituição, ela permite o uso da linguagem natural juntamente com a linguagem documentária? Quais os procedimentos?

11) Existem recomendações específicas para a indexação de documentos digitais? Se sim, qual os aspectos principais que ela recomenda?

12) A política de indexação, caso exista na instituição, determina uma quantidade de termos a serem utilizados na representação dos documentos? Se SIM, qual a média ela recomenda? Se NÃO, qual a média que você costuma utilizar?

13) TRADUÇÃO DE TERMOS PARA LINGUAGENS DOCUMENTÁRIAS

14) Você costuma levar em conta os termos utilizados pelos autores como palavras-chave no momento da representação documentária? Justifique.

15) Quais as partes do documento você percorre para identificar os termos de indexação antes de traduzi-los a partir da linguagem controlada? Explique como ocorre esse procedimento.

16) A tradução dos termos ocorre a partir da análise de assunto (leitura, identificação e seleção dos conceitos?)

17) Na sua opinião, qual a importância do uso de linguagens documentárias na tradução de termos para indexação de documentos digitais? Quais aspectos de sua utilização você considera mais importantes?

18) Na etapa de seleção e tradução de conceitos, você considera mais importante o aspecto organizacional ou a recuperabilidade proporcionados pelas linguagens documentárias?

19) Como é tratada a questão da especificidade no repositório?

20) Você concorda que o uso de termos qualificadores, também chamados de subdivisões, influenciam na recuperabilidade dos documentos, já que possibilitam uma indexação mais específica?

21) A exaustividade permite que todos os assuntos mais relevantes do documento sejam representados através linguagens documentárias. Existe orientações quanto a essa questão no repositório no qual você trabalha?

22) Uma pesquisa realizada através de um termo não autorizado em catálogos online de bibliotecas, retorna documentos indexados sob determinado assunto. Isto se deve ao fato de catálogos online usarem a estratégia de remissivas em seu sistema, o que não é possível ainda nos repositórios, podendo, assim, causar dispersão documental ou ocultação de documentos na pesquisa. Quais os procedimentos adotados por você para lidar com esta questão?

23) Os repositórios são ferramentas interoperáveis que formam uma rede de informação em âmbito mundial. Assim, no repositório que você trabalha, existe algum direcionamento para a tradução dos termos autorizados para outro idioma ou a tradução fica a cargo do próprio software (DSpace)?

24) Você já enfrentou resistência por parte dos autores quanto à tradução para uma linguagem documentária dos termos por eles utilizados? Se SIM, como solucionou este problema?

25) A construção e atualização das linguagens documentárias é baseada na garantia literária (termos mais citados nos textos científicos de determinada área do conhecimento) e isto pode implicar na ausência de termos indexadores para documentos produzidos em contextos específicos, como os culturais e geográficos. Dessa forma, você já sentiu dificuldades para indexar pesquisas relativas à região amazônica, da qual a Região Norte do Brasil faz parte, sendo que ainda são poucas e limitadas as linguagens documentárias voltadas especificamente para este local? Se SIM, como solucionou o problema?

26) Levando em consideração o rápido desenvolvimento do conhecimento científico e a necessidade de sua disseminação e recuperação, especialmente online, você considera que a atualização das linguagens documentárias acompanham esse desenvolvimento para indexação dos documentos? Justifique.