

**LUCIANA BEATRIZ PIOVEZAN RIO BRANCO**

**INTEROPERABILIDADE SEMÂNTICA ENTRE LINGUAGENS DE INDEXAÇÃO  
PARA BIBLIOTECAS UNIVERSITÁRIAS**

**Marília-SP  
2020**

**Programa de Pós-Graduação em Ciência da Informação**

**LUCIANA BEATRIZ PIOVEZAN RIO BRANCO**

**INTEROPERABILIDADE SEMÂNTICA ENTRE LINGUAGENS DE INDEXAÇÃO  
PARA BIBLIOTECAS UNIVERSITÁRIAS**

Tese apresentada ao Programa de Pós-Graduação em  
Ciência da Informação da Universidade Estadual  
Paulista como requisito para obtenção do título de  
Doutora.

**Orientadora:** Profa. Dra. Mariângela Spotti Lopes  
Fujita

**Coorientador:** Prof. Dr. Walter Moreira

**Área de concentração:** Informação, Tecnologia e  
Conhecimento

**Linha de pesquisa:** Produção e Organização da  
Informação

**Marília-SP  
2020**

R585i

Rio-Branco, Luciana Beatriz Piovezan

Interoperabilidade semântica entre linguagens de indexação para bibliotecas universitárias / Luciana Beatriz Piovezan Rio-Branco. -- Marília, 2020

142 f. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, Marília

Orientadora: Mariângela Spotti Lopes Fujita

Coorientador: Walter Moreira

1. Linguagem documentária. 2. Indexação. 3. Bibliotecas universitárias. 4. Tesouros. 5. Interoperabilidade semântica. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Filosofia e Ciências, Marília. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

**CERTIFICADO DE APROVAÇÃO**

TÍTULO DA TESE: INTEROPERABILIDADE SEMÂNTICA ENTRE LINGUAGENS DE INDEXAÇÃO PARA BIBLIOTECAS UNIVERSITÁRIAS

**AUTORA: LUCIANA BEATRIZ PIOVEZAN RIO BRANCO**

**ORIENTADORA: MARIANGELA SPOTTI LOPES FUJITA**

**COORIENTADOR: WALTER MOREIRA**

Aprovada como parte das exigências para obtenção do Título de Doutora em CIÊNCIA DA INFORMAÇÃO, área: Informação, Tecnologia e Conhecimento pela Comissão Examinadora:

Prof(a). Dr(a). MARIANGELA SPOTTI LOPES FUJITA  
Programa de Pós-Graduação em Ciência da Informação / Unesp, Faculdade de Filosofia e Ciências, Marília

Prof(a). Dr(a). DEISE MARIA ANTONIO SABBAG  
Programa de Pós-Graduação em Ciência da Informação / Universidade de São Paulo, Faculdade de Filosofia, Ciências e Letras, Ribeirão Preto

Prof(a). Dr(a). ISIDORO GIL LEIVA  
Gestión de la Información y de la Comunicación en las Organizaciones / Universidad de Murcia, Facultad de Comunicacion y Documentacion

Prof(a). Dr(a). BRIGIDA MARIA NOGUEIRA CERVANTES  
Departamento de Ciência da Informação / Universidade Estadual de Londrina

Prof(a). Dr(a). FRANCIELE MARQUES REDIGOLO  
Programa de Pós-Graduação em Ciência da Informação / Universidade Federal do Pará, Faculdade de Biblioteconomia

Marília, 21 de agosto de 2020



Prof(a). Dra. Marta Lígia Pomim Valentim  
Coordenadora do PPG em Ciência da Informação

*Ao meu irmão querido,  
com muitas saudades e a certeza de que te levarei comigo para sempre*

## AGRADECIMENTOS

À orientadora, *Profa. Dra. Mariângela Spotti Lopes Fujita*, pela confiança, incentivo, por ser muito mais do que uma orientadora, por ser quem confiou em mim, mesmo quando nem eu mais confiava, quem não desistiu de mim, mesmo nos momentos mais difíceis, e que ao fazer isso impediu que eu mesma desistisse de mim.

Ao coorientador *Prof. Walter Moreira* por todo apoio, por todos os cafés e caronas e pela oportunidade de realização do estágio de docência, em que aprendi muito mais do que imaginaria, mesmo minhas expectativas já sendo bem altas.

Aos membros da banca de qualificação, *Profa. Dra. Brígida Maria Nogueira Cervantes*, *Profa. Dra. Deise Antonio Sabbag*, agradeço pela atenção, paciência, sabedoria, solicitude e gentileza, pela compreensão e generosidade que demonstraram na realização da banca de qualificação deste trabalho, além dos importantes apontamentos oferecidos.

À *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)* pelo investimento realizado no financiamento desta pesquisa, com esperanças de que esta oportunidade continue a ser concedida às formiguinhas que trabalham noite e dia para construir o futuro da ciência brasileira. O presente trabalho foi realizado com apoio da *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001*.

Ao Grupo de Linguagem da Rede de Bibliotecas da Unesp.

Aos professores e funcionários do PPGCI Em especial gostaria de deixar minha gratidão à *Profa. Dulce Amélia de Brito Neves* e à *Profa. Paula Regina Dal'Evedove*, assim como estas, a todos os professores do Programa.

Ao *Felipe*, que sonhou meus sonhos, caminhou comigo e sem o qual nem sei onde estaria.

À minha amiga e comadre *Aline Farias*.

Aos companheiros dos desafios da Pós-Graduação, pelo exemplo de dedicação, graciosidade, ética e empatia, que provam diariamente que somos capazes de crescer conjuntamente. Agradeço a *Admeire Santos Sundström*, *Érica Fernanda Vitorini*, *Luciana Davanzo*, *Mona Cleide Quirino de Farias*, *Maria Carolina Andrade e Cruz*, *Roberta Vesú Alves*.

À comunidade da UNESP, pelos anos formadores e transformadores, pelos 12 anos de vivências, pela possibilidade de conhecer o *novo mundo* e pelo futuro que se abre.

Sou imensamente grata às inúmeras coincidências e alinhamentos que me permitiram trilhar este caminho, eu que fui a primeira da minha família a ingressar em uma faculdade, a única a cursar uma universidade pública, a primeira a seguir para a pós-graduação, a primeira a viajar de avião... E nada disso teria acontecido, nada que eu tenho e vivo de bom na minha vida hoje teria acontecido, assim como nenhum passo que darei ainda em minha vida seria possível, se não fosse pela existência da universidade pública e gratuita.

Minha sincera gratidão.

*“Nós tentamos fazer o melhor, mas veio o de sempre”*

*Viktor Stepanovich Chernomyrdin*



## RESUMO

A interoperabilidade semântica é a capacidade de sistemas de informação de trocar, usar e reusar informação compreensível por máquinas e se aplica às questões de construção e gerenciamento de linguagens de indexação para que estas possam ser utilizadas para indexação e recuperação de informação de forma integrada em diversos ambientes. Considerando este novo paradigma de representação do conhecimento, abordou-se como problema de pesquisa a disseminação de vocabulários com níveis de controle diversos construídos individualmente por bibliotecas universitárias brasileiras como fator de isolamento, por não atenderem às normativas internacionais e assim, não permitirem o alcance da interoperabilidade semântica com outros vocabulários. A hipótese levantada é que a interoperabilidade semântica tem importância estratégica para que bibliotecas universitárias possam construir e manter atualizadas linguagens de indexação de acordo com normativas internacionais necessárias à integração dos diferentes sistemas de representação de informação. Agregar interoperabilidade semântica às linguagens de indexação em bibliotecas universitárias, é a tese que garantirá a melhora nos processos de gestão destas linguagens, contribuindo para a sua atualização, a incorporação de termos, bem como para a percepção no aumento da qualidade da indexação realizada. Deste modo, a pesquisa teve por finalidade elencar as características preferidas no processo de construção de linguagens de indexação para que esta se constitua em uma linguagem interoperável, permitindo seu intercâmbio e reuso com outros sistemas de bibliotecas universitárias. Delineou-se como objetivo geral oferecer subsídios teóricos metodológicos à pesquisa sobre interoperabilidade semântica de linguagens de indexação no contexto teórico da organização e representação do conhecimento. E como objetivos específicos, realizar estudo teórico e metodológico sobre elaboração e manutenção de linguagens de indexação com foco na interoperabilidade semântica; analisar o processo de elaboração de linguagens de indexação sob o aspecto das normas internacionais focalizando a interoperabilidade semântica; identificar e analisar projetos de interoperabilidade semântica de linguagens de indexação e elaborar recomendações para a construção de linguagens com interoperabilidade semântica. Para alcançar tais objetivos, utilizou-se do método quadripolar de pesquisa, com uso de técnicas de análise de conteúdo e entrevistas semiestruturadas que foram realizadas com catalogadores de três bibliotecas universitárias e, posteriormente analisadas em categorias temáticas. Os resultados demonstraram que os projetos de linguagens de indexação com interoperabilidade semântica encontram-se em estágios diversos de desenvolvimento e estão sendo realizados sem o conhecimento das atuais normas de construção de linguagens de indexação. Foram identificadas dificuldades no desenvolvimento destes projetos que afetam as ações cotidianas de gestão da linguagem de indexação tendo efeitos no uso desta linguagem pelos usuários. Nesse sentido elaborou-se um conjunto inicial de recomendações, com base na ISO 25.964 partes 1 e 2, para contribuir com o desenvolvimento de projetos tais como os analisados. Conclui-se que o desconhecimento dos mapeamentos semânticos, enquanto elementos-chave indicados pelas normas internacionais para o estabelecimento de interoperabilidade semântica, advém da inacessibilidade destes documentos, por fatores principalmente idiomáticos e financeiros. A principal contribuição desta pesquisa foi demonstrar e analisar as propostas de mapeamentos semânticos e suas funções na interoperabilidade semântica entre linguagens de indexação.

**Palavras-Chave:** Linguagem de indexação; Indexação; Bibliotecas universitárias; Interoperabilidade semântica.

## ABSTRACT

Semantic interoperability is the ability of information systems to exchange, use and reuse information readable by machines and it applies to issues of building and managing indexing languages so that they can be used for information indexing and retrieval in an integrated way in various environments. Considering this new paradigm of knowledge representation, we had as a research problem the dissemination of vocabularies with different levels of control constructed individually by Brazilian university libraries as a factor of isolation, when they do not meet international standards and thus do not allow the semantic interoperability with other vocabularies. The hypothesis raised is that semantic interoperability has strategic importance so that university libraries can build and maintain updated indexing languages according to international guidelines necessary for the integration of the different information representation systems. Adding semantic interoperability to indexing languages in university libraries is the thesis that will ensure improvement in the management processes of these languages, contributing to their updating, the incorporation of terms, as well as the perception in increasing the quality of indexing performed. The purpose of this research was to identify the characteristics that should be preferred in the process of constructing indexing languages so that it constitutes an interoperable language, allowing its exchange and reuse with other systems of University libraries. We present as a general-purpose to offer theoretical and methodological contributions to research on semantic interoperability of indexing languages in the theoretical context of the knowledge organization. And as specific aims, we intended to carry out a theoretical and methodological study on the construction and maintenance of indexing languages with a focus on semantic interoperability; analyze the process of developing indexing languages under the aspect of international standards focusing on semantic interoperability; identify and analyze projects of semantic interoperability between indexing languages and present recommendations for the construction of indexing languages with semantic interoperability. To achieve these purposes, the Quadripolar method of research was used, employing techniques of content analysis and semi-structured interviews that were conducted with catalogers from three university libraries and afterward analyzed in thematic categories. The results demonstrated that the projects of indexing languages with semantic interoperability are in different stages of development and that they are being carried out without the knowledge of the current standards of construction of indexing languages. Difficulties in the development of these projects were identified that affect the daily actions of indexing language management having effects on the use of this language by users. In this sense, an initial set of recommendations was elaborated, based on ISO 25,964 parts 1 and 2, to contribute to the development of projects such as those analyzed. It is concluded that the obliviousness of semantic mappings, as key elements indicated by international standards for the establishment of semantic interoperability, results from the inaccessibility of these documents, due to mainly idiomatic and financial factors. The main contribution of this research was to demonstrate and analyze the proposals of semantic mappings and their functions in semantic interoperability between indexing languages.

**Keywords:** Indexing language; Indexing; University Libraries; Semantic interoperability.

## LISTA DE ILUSTRAÇÕES

Quadro 1 - Estrutura sistemática da pesquisa de acordo com seus objetivos.....	29
Quadro 2 - Análise conceitual – exaustividade .....	50
Quadro 3 - Análise conceitual – especificidade .....	53
Quadro 4 - Elementos legíveis por humanos usados para representar conceitos em mapeamentos .....	64
Quadro 5 - Tipos de mapeamentos encontrados na ISO 25.964-2 (2013) .....	66
Quadro 6 - Notações estabelecidas da ISO 25.964-2 (2013) para mapeamentos entre linguagens .....	67
Quadro 7 - Descrição dos procedimentos metodológicos e integração com polos da pesquisa .....	74
Quadro 8 - Sistematização da relação entre perguntas e categorias do protocolo de entrevistas .....	78
Quadro 9 - Comparação entre tópicos das normas ISO 25.964-1 (2011) e ISO 25.964-2 (2013) e as categorias de análise elaboradas .....	80
Quadro 10 - Sistematização da relação entre etapas da análise de conteúdo e categorias de análise .....	81
Figura 1 - Evolução das normas e padrões sobre tesouros .....	31
Figura 2 - Linha do tempo de marcos em normas e padrões para Tesouros no idioma inglês.	33
Figura 3 - Modelo 2 (link direto) como aplicado a quatro vocabulários.....	65
Figura 4 - Modelo 3 (modelo de hub) como aplicado a quatro vocabulários.....	66
Figura 5 - Sistematização do método quadripolar e a interação entre os polos.....	72

## LISTA DE ABREVIATURAS E SIGLAS

<b>ADL</b>	<i>Alexandria Digital Library</i>
<b>AIMS</b>	<i>Agriculture Information Management Standards</i>
<b>API</b>	<i>Application Programming Interface</i> (Interface de programação de aplicações)
<b>CGB</b>	Coordenadoria Geral de Bibliotecas
<b>COMAER</b>	Comando da Aeronáutica
<b>COSATI</b>	Comitê em Informação Científica e Técnica
<b>DCTA</b>	Departamento de Ciência e Tecnologia Aeroespacial
<b>DeCS</b>	Descritores em Ciências da Saúde
<b>FAO-UN</b>	<i>Food and Agriculture Organization of the United Nations</i>
<b>FAPESP</b>	Fundação de Amparo à Pesquisa do Estado de São Paulo
<b>FRAD</b>	<i>Functional Requirements for Authority Data</i> (Requisitos Funcionais para Dados de Autoridade)
<b>FRBR</b>	<i>Functional Requirements for Bibliographic Records</i> (Requisitos Funcionais para Registros Bibliográficos)
<b>FRSAD</b>	<i>Functional Requirements for Subject Authority Data</i> (Requisitos Funcionais para Dados de Autoridade Assunto)
<b>IA-DOC</b>	Biblioteca da Divisão de Informação e Documentação
<b>IFLA</b>	<i>International Federation of Library Associations and Institutions</i> (Federação Internacional de Associações e Instituições Bibliotecárias)
<b>ITA</b>	Instituto de Tecnologia da Aeronáutica
<b>FCLAr</b>	Faculdade de Ciências e Letras de Araraquara
<b>GARE</b>	<i>Guidelines for Authority and References Entries</i> (Diretrizes para Autoridades e Entradas de Referência)
<b>GARR</b>	<i>Guidelines for Authority Records and References</i> (Diretrizes para Registros de Autoridade e Referência)
<b>LC</b>	<i>Library of Congress</i> (Biblioteca do Congresso dos Estados Unidos da América)
<b>LCSH</b>	<i>Library of Congress Subject Headings</i>
<b>LEM BP</b>	<i>Lista de encabezamientos de materia para Bibliotecas Públicas del Ministerio</i>
<b>MADS</b>	<i>Metadata Authority Description Schema</i> (Esquema de Descrição de Autoridades em Metadados)
<b>MARC21</b>	<i>MACHine-Readable Cataloguing 21<sup>st</sup> edition</i>

<b>MASP</b>	Museu de Arte de São Paulo
<b>MeSH</b>	<i>Medical Subject Headings</i>
<b>OPAC</b>	<i>Online Public Access Catalogs</i> (Catálogo Público de Acesso Online)
<b>PRECIS</b>	PREserved Context Index System
<b>RDF</b>	<i>Resource Description Framework</i>
<b>SIBi/USP</b>	Sistema Integrado de Bibliotecas da USP
<b>STATI</b>	Seção Técnica de Aquisição e Tratamento da Informação
<b>STRAUD</b>	Seção Técnica de Referência, Atendimento ao Usuário e Documentação
<b>SKOS</b>	<i>Simple Knowledge Organization Systems</i>
<b>SOC</b>	Sistema de Organização do Conhecimento
<b>TBN</b>	Terminologia da Biblioteca Nacional
<b>TEST</b>	Tesouro de Engenharia e Termos Científicos
<b>TIC</b>	Tecnologias de Informação e Comunicação
<b>URI</b>	<i>Uniform Resource Identifier</i> (Identificador Universal de Recursos)
<b>UNIMARC</b>	<i>United International Machine-Readable Cataloging</i>
<b>UNISIST</b>	<i>United Nations Educational, Scientific and Cultural Organization</i>
<b>Unifesp</b>	Universidade Federal de São Paulo
<b>Unesp</b>	Universidade Estadual Paulista “Júlio de Mesquita Filho”
<b>VUSP</b>	Vocabulário Controlado do SIBi/USP
<b>W3C</b>	<i>World Wide Web Consortium</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>14</b>
<b>1.1</b>	<b>Delimitação do problema e tese .....</b>	<b>21</b>
<b>1.2</b>	<b>Objetivos .....</b>	<b>25</b>
<b>1.3</b>	<b>Justificativa .....</b>	<b>26</b>
<b>1.4</b>	<b>Sistemática da tese .....</b>	<b>27</b>
<b>2</b>	<b>AS LINGUAGENS DE INDEXAÇÃO: ASPECTOS TEÓRICOS E NORMATIVOS .....</b>	<b>30</b>
<b>2.1</b>	<b>Normas internacionais de construção de linguagens de indexação.....</b>	<b>30</b>
<b>2.2</b>	<b>Características dos tesouros e das linguagens de indexação na perspectiva da norma ISO 25.964-1 .....</b>	<b>34</b>
<b>2.3</b>	<b>Escolha da Linguagem de indexação: construção de uma linguagem própria ou adoção de uma linguagem existente .....</b>	<b>44</b>
<b>3</b>	<b>INTEROPERABILIDADE SEMÂNTICA DE LINGUAGENS DE INDEXAÇÃO: PANORAMA DAS NORMAS E PADRÕES INTERNACIONAIS .....</b>	<b>55</b>
<b>3.1</b>	<b>Modelos de mapeamento semântico.....</b>	<b>63</b>
<b>3.2</b>	<b>Formatos e padrões de metadados para dados de autoridade .....</b>	<b>67</b>
<b>4</b>	<b>METODOLOGIA.....</b>	<b>72</b>
<b>4.1</b>	<b>O polo técnico .....</b>	<b>74</b>
<i>4.1.1</i>	<i>Elaboração do protocolo de entrevista e realização de entrevista semiestruturada ....</i>	<i>77</i>
<i>4.1.2</i>	<i>Universo de pesquisa.....</i>	<i>81</i>
<b>4.2</b>	<b>O polo morfológico .....</b>	<b>84</b>
<b>5</b>	<b>APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS .....</b>	<b>85</b>
<b>5.1</b>	<b>Análise das entrevistas com os procedimentos de Análise de Conteúdo .....</b>	<b>85</b>
<b>5.2</b>	<b>Síntese dos resultados: recomendações para construção e manutenção de linguagens de indexação visando a interoperabilidade semântica .....</b>	<b>98</b>
<b>6</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>100</b>
	<b>REFERÊNCIAS.....</b>	<b>105</b>
	<b>APÊNDICE A – TRANSCRIÇÃO DE ENTREVISTA I.....</b>	<b>114</b>
	<b>APÊNDICE B - TRANSCRIÇÃO DE ENTREVISTA II.....</b>	<b>119</b>
	<b>APÊNDICE C - TRANSCRIÇÃO DE ENTREVISTA III .....</b>	<b>130</b>

<b>APÊNDICE D – QUADRO DE ANÁLISE DAS TRANSCRIÇÕES - ETAPA DE PRÉ-ANÁLISE .....</b>	<b>133</b>
<b>APÊNDICE E – QUADRO DE ANÁLISE DAS TRANSCRIÇÕES - ETAPA DE EXPLORAÇÃO DO MATERIAL.....</b>	<b>139</b>

## 1 INTRODUÇÃO

Apresentamos pesquisa que tem como temática a interoperabilidade semântica na construção de linguagens de indexação para representação e recuperação do conhecimento em bibliotecas universitárias. Deste modo, se insere na área da Ciência da Informação, mais especificamente na linha de Produção e Organização da Informação.

Em Ciência da Informação, é a Organização do Conhecimento a área que se ocupa de processos e instrumentos para o tratamento da informação documental visando o aperfeiçoamento da recuperação da informação em ambientes de informação. Segundo Fujita (2001), a Organização do Conhecimento tem suas raízes ligadas à criação de sistemas de classificação bibliográfica para organização e representação temática de coleções documentais.

Os processos referentes à Organização do Conhecimento – análise, representação, armazenamento e recuperação da informação, têm por finalidade o uso da informação, individual e social. Trabalhamos nessa área com a perspectiva de que se não houvesse tais processos a informação, enquanto conhecimento social, provavelmente perder-se-ia no tempo.

Em um momento social em que são sentidas constantes evoluções no campo das Tecnologias de Informação e Comunicação (TIC) e com isso também um acréscimo de sua utilização nas bibliotecas universitárias volta-se a preocupação ao acesso à informação disponibilizada em tais ambientes.

A disseminação dos catálogos online, dos repositórios institucionais e a vinculação de objetos digitais a estes, representa um diferencial aos serviços de informação oferecidos nas bibliotecas universitárias, porém não é desejável oferecer aos usuários ferramentas de pesquisa de difícil acesso, especialmente diante da grande disseminação de ferramentas de busca de fácil utilização, como a plataforma *Google*. Dessa forma, parece ilusório esperar que os usuários de bibliotecas universitárias se adaptem às particularidades de cada catálogo e repositório que têm a sua disposição, realizando buscas individuais em cada um destes ambientes e conhecendo cada um dos vocabulários empregados na indexação de seus documentos.

Para esta questão, o tema das linguagens de indexação é central. O uso das linguagens de indexação tem por finalidade uniformizar a representação dos conceitos expressivos dos assuntos de documentos, em razão da diversidade conceitual existente entre o discurso dos autores, indexadores e usuários. Seu papel mediador é de suma importância para a recuperação da informação, atuando tanto na representação do indexador, no momento da indexação, quanto na formulação de estratégias de busca por parte dos usuários.



As reflexões apresentadas aqui são derivadas de uma trajetória de pesquisa que iniciou-se durante os anos de Graduação em Biblioteconomia onde, por meio de participação em projeto de Iniciação Científica desenvolvido com tema Política de indexação, no qual foi utilizada a metodologia de observação participante com aplicação de questionários de diagnóstico e realização de visitas para observação participante da prática de indexação e identificação de elementos de política de indexação em bibliotecas universitárias da região de Marília-SP (SANTOS, 2011).

Durante a realização desta pesquisa identificou-se que o processo de indexação é realizado por bibliotecários capacitados por Graduação em Biblioteconomia, porém que nem sempre seguem princípios equânimes claramente estabelecidos, ou que tenham sido registrados em um documento, o que pode gerar disparidades nos resultados da indexação afetando a qualidade do processo, bem como, da recuperação da informação.

Os resultados desta investigação demonstraram também que o processo de indexação realizado nestas bibliotecas universitárias não tem por hábito receber avaliações de modo a verificar sua eficiência e qualidade e, portanto, a eventual necessidade de mudança nas práticas adotadas.

Em pesquisa de mestrado abordamos a indexação e avaliação da indexação em bibliotecas universitárias para recuperação temática (PIOVEZAN, 2015). Nesta pesquisa foi desenvolvido um estudo de avaliação da indexação em catálogos online de dois sistemas de bibliotecas universitárias com utilização do método de avaliação extrínseca mediante a interconsistência, que consiste em avaliar os resultados da indexação de um mesmo documento realizada por dois ou mais indexadores, ou sistemas de indexação ou ainda indexadores humanos e automáticos, de modo a observar se há consistência nos termos atribuídos. Esse método de avaliação permite analisar os aspectos da indexação que lhe conferem qualidade, posto que ao observar a consistência, torna-se possível também analisar a exaustividade e a precisão da indexação de forma contextualizada.

A proposta deste estudo foi, portanto, realizar avaliação comparada da indexação, de modo a observar os elementos de política de indexação que têm por finalidade determinar a qualidade da indexação, com o objetivo de analisar a consistência enquanto indicador de qualidade da indexação em catálogos online de dois sistemas de bibliotecas universitárias. Para atingir esta finalidade, foi aplicado o método de avaliação extrínseca mediante a interconsistência (PIOVEZAN, 2015).

Esta avaliação, cuja análise teve embasamento também nos manuais de indexação de ambos os sistemas de bibliotecas universitárias, demonstrou que os níveis de consistência na

indexação entre esses dois catálogos ficaram abaixo da média esperada de acordo com a literatura (GIL LEIVA, 2001), bem como demonstrou o efeito da exaustividade e da precisão da indexação nesses índices.

Pode-se observar também, que a consistência se mostrou um bom indicador para análise da precisão e da especificidade da indexação visto que demonstra o uso da linguagem no contexto legítimo de indexação, permitindo avaliar, por exemplo, a necessidade de atualização e ampliação da linguagem de indexação adotada pelo sistema, bem como, a relação entre a especificidade pretendida e a especificidade real alcançada pelo sistema (PIOVEZAN, 2015).

Diante dos resultados e das considerações suscitadas por essa avaliação percebeu-se que a consistência da indexação é fortemente sustentada pela linguagem de indexação utilizada para traduzir os conceitos identificados nos documentos. Para alcançar níveis mais elevados de consistência da indexação e assim melhorar a qualidade do processo em bibliotecas universitárias é desejável que seja utilizada uma linguagem única na indexação, e que esta linguagem esteja atualizada e atenda aos critérios de especificidade e exaustividade compatíveis com a robustez de um sistema de bibliotecas universitárias.

Nesse sentido, a adoção de uma linguagem compartilhada, como aquelas construídas por instituições responsáveis por serviços cooperativos, desponta como uma solução para que as bibliotecas tenham à sua disposição uma linguagem ampla e constantemente atualizada.

Em âmbito internacional, verificam-se experiências de uso compartilhado de linguagens, tais como a *Lista de encabezamientos de materia para Bibliotecas Públicas del Ministerio* – LEM BP, utilizada pelas bibliotecas públicas da Espanha; a *Library of Congress Subject Headings* – LCSH, desenvolvida pela Biblioteca do Congresso dos Estados Unidos da América; e a *Medical Subject Headings* – MeSH, da *National Library of Medicine*, dos Estados Unidos da América.

No Brasil, verifica-se, no âmbito das bibliotecas universitárias, o uso difundido das linguagens LCSH e MeSH, como observado por Fujita e Santos (2016b). Já como experiências nacionais de construção de linguagens compartilhadas encontram-se a Lista de Cabeçalhos de Assunto da Rede BIBLIODATA e a Terminologia de Assuntos da Biblioteca Nacional, ambas derivadas em sua matriz estrutural da LCSH, além do Vocabulário Controlado do SIBi/USP – VUSP construído pela Universidade de São Paulo para uso próprio e que pode também ser utilizado por outras instituições (FUJITA; SANTOS, 2016a).

As linguagens de indexação são instrumentos de controle de vocabulário que objetivam a representação e a recuperação de informação, utilizados na indexação, para explicitar o

conteúdo de documentos e representá-los em sistemas de recuperação da informação, e na expressão das necessidades de pesquisa de usuários, para formulação de estratégias de busca.

A construção das linguagens de indexação tem por fundamento os princípios teóricos e metodológicos provenientes da literatura sobre Linguística, Terminologia, Teoria do Conceito, Teoria da Classificação e Recuperação da Informação. E por objetivos o controle de vocabulário, no que tange à eliminação da ambiguidade, controle de sinonímia presentes na linguagem natural e a normalização das variações de gênero e número; oferecer auxílio ao indexador na seleção de termos que melhor representem o assunto dos documentos; a mediação entre o conhecimento armazenado em sistemas e os usuários e facilitar a seleção de termos para formulação de estratégias de busca.

Existe em Organização do Conhecimento uma diversidade de instrumentos para representação do conhecimento e uma diversidade terminológica ao nomear e classificar tais instrumentos. Como observado em Fujita, Santos e Alves (2018) identificam-se os instrumentos que compreendem os sistemas de classificação, taxonomias, tesouros, listas de cabeçalhos de assunto, bem como, as ontologias que surgiram mais recentemente no escopo deste domínio.

Da mesma maneira, são encontrados na literatura de Organização do Conhecimento o uso dos termos vocabulário controlado, linguagem documentária, linguagem de indexação e, mais recentemente, sistemas de organização do conhecimento, para designar as tipologias destes diversos instrumentos de representação do conhecimento.

É interessante observar a colocação de Moreira (2019) sobre este tipo de situação, que explica que “a utilização de determinados termos em detrimento de outros implica a opção, de modo consciente ou não, por um determinado campo conceitual e pelo conjunto terminológico que lhe caracteriza” (MOREIRA, 2019, p. 16). Sendo assim, observa-se que os termos mencionados filiam-se, em um primeiro momento, às diferentes correntes de pensamento coexistentes em Organização do Conhecimento.

Desse modo, para descrever os instrumentos utilizados para os processos de classificação e indexação têm-se a adoção, do termo linguagem documentária pela linha teórica francesa, tradicionalmente denominada análise documentária; o uso do termo linguagem de indexação pela linha teórica anglo-saxã, mais especificamente de tradição inglesa – denominada indexação e em menor escala na tradição norte-americana, denominada catalogação de assunto.

Fujita, Santos e Alves (2018) em estudo sobre este fenômeno de variação terminológica referente aos instrumentos e, especialmente, às categorias de instrumentos de representação do conhecimento verificaram que, por outro lado, também existe uma perspectiva pragmática na adoção do termo, identificada especialmente quando o conceito escolhido é sistemas de

organização do conhecimento, isto porque a comunidade de pesquisa na área de Organização do Conhecimento reconheceu a necessidade de adotar um termo que compreendesse também, e em certa medida inclusive com maior ênfase, os instrumentos voltados para a Organização do Conhecimento com funções de representação e recuperação do conhecimento em ambientes digitais, bem como, no âmbito da web semântica.

A busca no contexto da Organização do Conhecimento por uma caracterização genérica para estes instrumentos faz parte da compreensão de que são instrumentos que se assemelham em suas características e que, principalmente, possuem um objetivo principal em comum, a representação do conteúdo de documentos que facilite a sua recuperação (BARITÉ, 2011).

Nessa compreensão observada por Fujita, Santos e Alves (2018) em sua análise, os termos linguagem de indexação, linguagem documentária e sistemas de organização do conhecimento (SOC) são considerados como termos equivalentes com o propósito definido de desempenhar funções de controle de vocabulário, embasando os processos de Organização e Representação do Conhecimento.

Nesse sentido, Golub (2011, p. 1), aponta o termo vocabulário controlado para identificar qualquer conjunto ou lista de termos controlados que tenha como função ser utilizado na descrição de metadados de documentos. A própria autora indica, como objeto de comparação, as linguagens de indexação que também são em sua visão tipos de vocabulários controlados, porém mais formalizadas e utilizadas especificamente para descrever documentos quanto ao seu assunto objetivando a recuperação da informação. Assim, em bibliotecas são utilizadas linguagens de indexação que possuem controle de vocabulário de tipologias tais como os vocabulários controlados, as listas de cabeçalhos de assunto e os tesouros.

Diante destas considerações, nesta pesquisa será usado o termo linguagem de indexação que se compreende como termo genérico que agrupa os instrumentos de representação do conteúdo temático de documentos em ambientes de Organização do Conhecimento, em especial as bibliotecas universitárias.

Para que uma linguagem de indexação seja utilizada por mais de uma instituição para a indexação de seus documentos e também permita a integração de diferentes sistemas – catálogos online, bases de dados, repositórios temáticos e institucionais – para a representação e recuperação integrada da informação é preciso que esteja adequada aos padrões atuais de interoperabilidade semântica.

A interoperabilidade, de modo geral, é definida pela norma ISO 25.964-2:2013 como a habilidade de dois ou mais sistemas ou componentes de trocar informação e de usar a informação que foi trocada. Embora o termo interoperabilidade seja carregado de significações

relativas às tecnologias mais atuais, o conceito atrelado a este termo é bastante conhecido nas discussões da Biblioteconomia. Exemplo mais destacado deste fato é o desenvolvimento do Formato MARC para padronização e intercâmbio de dados bibliográficos e sua massiva utilização por bibliotecas e centros de informação a partir da década de 1960 até os dias atuais. De modo a permitir esta troca de informações também foi criado o protocolo ANSI/NISO Z39.50, cuja primeira versão foi lançada em 1988, e a norma ISO 2709, datada em sua primeira versão de 1981, que oferecem diretrizes para a estrutura de um registro bibliográfico e a comunicação de máquinas para pesquisa e recuperação de informação.

Andrade (2012) explica que a interoperabilidade pode ser compreendida em níveis específicos, ou tipologias, que se referem aos diversos aspectos envolvidos em sua constituição. Com base em Sayão e Marcondes (2008), Andrade (2012) elenca como níveis de interoperabilidade a interoperabilidade técnica, que é referente aos protocolos e formatos de metadados e “[...] permite que os sistemas e os hardwares se reconheçam” (ANDRADE, 2012, p. 119); a interoperabilidade semântica, referente aos significados das informações compartilhadas; a interoperabilidade política/humana, que se relaciona com as políticas de informação adotadas; a interoperabilidade intercomunitária, que se refere às questões de interdisciplinaridade; a interoperabilidade legal, que está relacionada aos aspectos legais de acesso à informação e a interoperabilidade internacional, que se refere às questões de cooperação internacional, barreiras linguísticas e às normas e padrões internacionais.

Os diversos ambientes de informação, como os catálogos online, os repositórios institucionais, as bases de dados, e os diversos objetos digitais a elas vinculados, demandam linguagens de indexação complexas que atendam a esse novo paradigma de representação do conhecimento, permitindo o tratamento da informação em suas estruturas semânticas. Tais linguagens de indexação exigem não apenas tecnologias adequadas à sua construção e gerenciamento, mas também normas e metodologias apropriadas, que facilitem sua construção, manutenção e uso, e que também permitam sua maior inserção e intercâmbio no ambiente digital ao estarem padronizadas.

Em ambientes de informação a principal meta que recai sobre a interoperabilidade é de “[...] tentar unificar a representação, manipulação ou a transmissão de algum item de informação de forma que dois ou mais diferentes sistemas possam ‘compreender’ igualmente este item” (SAYÃO, 2007, p. 20). Sayão (2007) coloca que a interoperabilidade deve ser um objetivo continuamente perseguido, posto que está relacionada não apenas às questões tecnológicas, mas também culturais e políticas, envolvendo não apenas sistemas, como também procedimentos, relacionamentos e leis.

Moreira e Lara (2012) lembram que para que sistemas de informação sejam capazes de realizar essa cooperação no intercâmbio e reuso de informações é preciso que tenham sido construídos visando esta filosofia de cooperação. Porém, mais recentemente, com a publicação do *Simple Knowledge Organization Systems* (SKOS) pelo consórcio *World Wide Web Consortium* (W3C) que se integra ao conjunto de softwares de gerenciamento de linguagens de indexação, também é possível adaptar as linguagens existentes no intuito de agregar interoperabilidade semântica. O padrão SKOS será abordado no capítulo 3 deste trabalho.

No contexto dos catálogos de bibliotecas e demais bases de dados bibliográficos a interoperabilidade atua na estrutura dos metadados utilizados para a descrição dos documentos, assim como nas linguagens de indexação utilizadas para representar o conteúdo destes documentos. A interoperabilidade no âmbito do tratamento de assunto se aplica, portanto, às questões de construção e gerenciamento de linguagens de indexação, sendo de ordem semântica.

Compreende-se a interoperabilidade semântica para os fins desta pesquisa como:

[...] a capacidade dos sistemas de informação (bibliotecas digitais e repositórios institucionais) para intercambiar informações baseando-se em um significado comum dos termos e expressões contidas nos metadados e documentos, com o objetivo de assegurar a consistência, a representação e a recuperação dos conteúdos. Isto envolve o uso de esquemas semânticos que incluem vocabulários controlados [...] (GÓMEZ DUEÑAS<sup>1</sup>, 2012, p. 3, traduzido por ANDRADE; CERVANTES, 2012, p 154).

O tema interoperabilidade semântica tem sido objeto de estudo de pesquisadores em nível global (MAI, 2003; MARTINEZ TAMAYO et al., 2011; ANDRADE, 2012; ANDRADE, 2015; EL HADI, 2015; LUCARELLI; VITI, 2015; SOERGEL, 2015; ANDRADE; LARA, 2016; ARAVE; JACOB, 2016; GARCIA MARCO, 2016; MAYR et al., 2016; DU et al., 2017; LLANES PADRÓN; PASTOR SÁNCHEZ, 2017; ANDRADE, 2018) tornando-se uma questão de vital interesse à agenda de pesquisa em Organização do Conhecimento ao objetivar a estruturação da informação para que seja possível trocar, usar e reusar não apenas dados, mas conteúdos contextualizados semanticamente, ou seja, informação dotada de significado compreensível por máquinas.

Dessa forma o foco da interoperabilidade semântica reside na estruturação da informação (atribuição dos metadados de documentos) por meio do controle de vocabulário, ou

---

<sup>1</sup> GÓMEZ DUEÑAS, Laureano Felipe. Modelos de interoperabilidad en bibliotecas digitales y repositorios documentales: caso Biblioteca Digital Colombiana BDCOL.

seja, com o uso de linguagens estruturadas constituídas com base na padronização nos níveis tradicionais (relações sintáticas e semânticas) assim como em nível “inter-linguagens” inventariando também as relações semânticas entre conceitos presentes em mais de uma linguagem (MARTINS, 2013).

A interoperabilidade semântica não é algo novo, o conceito encontra lastro no âmbito da Biblioteconomia e na Ciência da Informação nos conceitos de compatibilidade e conversibilidade presentes na literatura da área desde a década de 1960. A compatibilidade foi definida por Newman (1965) como “a capacidade de um sistema de informação aceitar dados da indexação e resumo de outro sistema sobre qualquer outro assunto que seja comum a ambos” (NEWMAN, 1965<sup>2</sup>, p. 7 apud ANDRADE, 2015, p. 70). Já a conversibilidade pode ser compreendida com base na explicação de Dahlberg para o processo de conversão como sendo “o processo de transformar registros de informação, baseado na transcrição de códigos, estruturas de dados, etc., fazendo com que os mesmos sejam intercambiáveis entre dois ou mais sistemas que utilizam diferentes convenções e mídias” (DAHLBERG, 1981).

A compatibilidade e conversibilidade têm sido compreendidas enquanto conceitualmente equivalentes e que se diferenciam da interoperabilidade semântica ao proporem a realização da integração de forma “manual”, pelos métodos de “léxico intermediário” (NEVILLE, 1972) e de “matriz de compatibilidade conceitual” (DAHLBERG, 1981).

A iniciativa de agregar interoperabilidade aos sistemas de recuperação de informação visa ao “objetivo de criar condições para que os usuários possam acessar múltiplas bases de dados, usando uma interface definida e uma linguagem de busca normalizada” (MOREIRA; LARA, 2012, p. 2). Para tanto, a interoperabilidade das linguagens de indexação demanda compatibilidade nos níveis técnico, estrutural e conceitual (BOCCATO; TORQUETTI, 2012).

### **1.1 Delimitação do problema e tese**

Entende-se que as questões tratadas no âmbito das discussões sobre compatibilidade e conversibilidade são abrangidas e complementadas pela pesquisa sobre interoperabilidade semântica, o que determina que o foco de estudos desta pesquisa deva estar sobre a interoperabilidade semântica entre linguagens de indexação.

---

<sup>2</sup> NEWMAN, S. M. Information systems compatibility. Washington [D.C.]: Spartan Books; London [Engl.]: Macmillan, 1965.

Gil Leiva (2008) indica a consideração de dois fatores que influenciam a interoperabilidade quando se trata de linguagens de indexação: a) ser fundamental a similaridade temática das diferentes linguagens; b) a compatibilização de linguagens de indexação sobre o mesmo assunto e no mesmo idioma pode economizar tempo e dinheiro.

A capacidade de interconexão com outros sistemas, realizando intercâmbio de dados com a manutenção de sua coesão se coloca como um indicador de qualidade de linguagens de indexação no contexto atual (PASTOR SANCHEZ; MARTINEZ MÉNDEZ; RODRIGUEZ MUNHOZ, 2012). Porém, para que se alcance um instrumento consistente é preciso estar atento às diretrizes apresentadas pelas normas internacionais já mencionadas.

É preciso também estar ciente de que as linguagens de indexação são instrumentos vivos que devem estar em constante atualização, mimetizando o processo vivenciado pelas linguagens naturais. Fujita (2011) ainda lembra que a tarefa de construir linguagens para a indexação

[...] é complexa para bibliotecários que, em sua atuação profissional em bibliotecas ou em sistemas de bibliotecas de áreas de especialidade, necessitam oferecer linguagens documentárias de especialidade permanentemente atualizadas para a indexação e recuperação da informação documentária (FUJITA, 2011, p. 37).

Este processo de construção e gerenciamento de linguagens é bastante complexo, posto que para atender as necessidades de Organização do Conhecimento em bibliotecas universitárias a linguagem utilizada deve constituir-se de vocabulário estruturado, abrangente a todas as áreas do conhecimento e dotado de exaustividade e especificidade.

No contexto das Instituições de Ensino Superior brasileiras esse desafio se complexifica também pela necessidade de uma linguagem compatível não apenas com outras linguagens nacionais, mas também com linguagens construídas em outros idiomas, visando o objetivo de internacionalização dessas instituições. De modo a permitir melhor acesso ao sistema por usuários falantes de outros idiomas e dando maior visibilidade à produção científica nacional.

As bibliotecas universitárias representam a vanguarda das atividades de Organização do Conhecimento e nas bibliotecas universitárias, a linguagem adotada deverá, necessariamente, estar atualizada e completa quanto a estes quesitos abordados. A interoperabilidade semântica é estratégica, portanto, ao abrir a possibilidade de uso de linguagem completa de atualização facilitada, que permita a integração de termos e conceitos. Essa relevância outorgada à interoperabilidade semântica se torna clara ao se observar a dispersão existente em bibliotecas



brasileiras, que vem buscando construir linguagens de indexação próprias para uso em seus catálogos.

Em pesquisa sobre o uso destas linguagens em bibliotecas universitárias do sudeste brasileiro Fujita et. al (2019) identificou que das 60 bibliotecas pesquisadas 49 utilizam linguagem de indexação em seus catálogos, sendo que 24 utilizam mais de uma linguagem de indexação e outras 14 utilizam a combinação de linguagem de indexação e linguagem natural no processo de indexação. Em uma tentativa de fugir do uso de linguagem natural, 20 das bibliotecas pesquisadas relataram utilizar linguagem de indexação de elaboração própria e mais 10 afirmaram que pretendem elaborar suas próprias linguagens.

Porém, o mesmo estudo observou que estas bibliotecas que possuem ou pretendem elaborar suas próprias linguagens, não dominam as ferramentas necessárias para tal esforço. Por exemplo, das 30 bibliotecas nesta situação apenas uma citou o uso de um software específico de gestão de linguagens para a tarefa (FUJITA et al., 2019). Este dado é relevante pois destaca um ponto fraco na elaboração destes vocabulários, visto que o processo já complexo de construção adquire um aspecto de maior dificuldade ao ser feito sem o apoio automático e direcionado por softwares desenvolvidos para esta tarefa.

Já no ano de 2005 Currás (p. 219) prevenia que a construção manual não era mais concebível a não ser em casos muito particulares e cujo universo de termos e documentos para trabalho fosse nas palavras da autora “mínimo”. Currás (2005) inclusive já colocava a existência de diversos softwares com as funcionalidades requeridas e que variavam amplamente quanto ao investimento necessário, de modo que mesmo pequenas instituições poderiam considerar estas opções. Existem, ademais, os softwares gratuitos desenvolvidos muitas vezes por pesquisadores inseridos em Organização do Conhecimento. Como exemplos destes, tem-se o software Tematres, desenvolvido pelo pesquisador argentino Diego Ferreyra, que é um software livre e “[...] seu sistema também permite desenvolver estruturas de navegação web e funcionar como complemento de um sistema de gerenciamento de bibliotecas, para a gestão da linguagem” (FUJITA et al., 2017, p. 69). E também o software Vocbench, desenvolvido pelo *Agriculture Information Management Standards (AIMS)* da *Food and Agriculture Organization of the United Nations (FAO-UN)* inicialmente para uso no gerenciamento colaborativo do tesouro Agrovoc e posteriormente disponibilizado gratuitamente, trata-se de uma “[...] plataforma colaborativa para o gerenciamento de tesouro, disponível gratuitamente e de código aberto, oferecendo suporte nativo de RDF para sistemas de organização de conhecimento em SKOS e SKOS-XL” (STELLATO et al., 2017, p. 2, tradução livre).

Fica claro que os diversos vocabulários que vêm sendo desenvolvidos de forma isolada e, a partir do estudo de Fujita et al. (2019) pode-se inferir que, sem procedimentos criteriosos de elaboração - tais como, o atendimento à normalização internacional vigente, controle de vocabulário rigoroso e o uso de softwares apropriados, se colocarão como impedimentos à integração de sistemas de representação da informação no contexto das bibliotecas universitárias brasileiras.

O compartilhamento de linguagens de indexação demanda negociação e aceitação para o intercâmbio consistente de dados. Esta constatação nos leva a crer que uma pesquisa que objetive atuar no tema da interoperabilidade no âmbito da organização do conhecimento deva se desenvolver com foco não apenas nos níveis técnicos e tecnológicos do tema, mas também no que diz respeito ao elemento humano, realizando análises qualitativas que ponderem os aspectos político, contextual e social envolvidos.

Com base na contextualização do problema realizada até aqui, destacamos as três premissas principais nas quais nossa pesquisa se baseia

- A interoperabilidade semântica não é algo novo, porém após mais de 50 anos de aparente de lentidão de desenvolvimento no campo da Organização da Conhecimento, o tema encontra atualmente possibilidade de efetiva evolução diante do interesse crescente no debate teórico sobre suas vantagens e meios para alcançá-la, exemplificado com a publicação da norma internacional ISO 25.964, e do desenvolvimento de tecnologias computacionais que possam acompanhar os debates teóricos (vide página 21);
- As bibliotecas universitárias representam a vanguarda das atividades de Organização do Conhecimento no contexto brasileiro, sendo centros que concentram e impulsionam a existência de diversos serviços de informação, como os catálogos online, os repositórios institucionais e as bases de dados, que incorporam uma amplitude de objetos educacionais digitais, proporcionando acesso ao conhecimento especializado. O acervo de conhecimento das universidades, portanto, está sob responsabilidade de suas bibliotecas, porém muitas vezes a universidade também possui arquivos e museus que custodiam parte deste conhecimento, de modo que a adoção de tecnologias que permitam integrar estes ambientes informacionais é essencial ao melhor desempenho das universidades (vide página 23)
- Os diversos vocabulários que vêm sendo desenvolvidos de forma isolada e sem procedimentos criteriosos de elaboração, seja a partir da combinação imponderada de mais de uma linguagem ou do uso da linguagem natural para completar

inconsistências da linguagem adotada, seja construindo linguagens de indexação próprias no intuito de indexar seus catálogos sem que para isso tenham domínio das ferramentas necessárias para o desenvolvimento destes processos, como o uso de softwares apropriados, bem como sem o apoio das atuais diretrizes normativas acerca da construção de linguagens de indexação representam obstáculos ao alcance da interoperabilidade semântica e, por conseguinte, à integração dos sistemas de representação e recuperação de informação para bibliotecas universitárias (vide página 20).

Movida por estas premissas, a pesquisa norteia-se a partir do questionamento/problema: A disseminação de vocabulários com níveis de controle diversos construídos individualmente por bibliotecas universitárias brasileiras como fator de isolamento, por não atenderem às normativas internacionais e assim não permitirem o alcance da interoperabilidade semântica com outros vocabulários.

A hipótese que se procura corroborar é a de que:

Para que bibliotecas universitárias possam construir e manter atualizadas linguagens de indexação que atendam às normativas internacionais necessárias à integração dos diferentes sistemas de representação de informação a interoperabilidade semântica tem importância estratégica.

Desse modo, tem como tese de pesquisa:

Agregar interoperabilidade semântica às linguagens de indexação em bibliotecas universitárias, garantirá a melhora nos processos de gestão destas linguagens, contribuindo para a sua atualização, a incorporação de termos, bem como para a percepção no aumento da qualidade da indexação realizada.

Para tanto, propõe-se estudo teórico e metodológico sobre a interoperabilidade semântica entre linguagens de indexação no contexto da política de indexação de modo a elencar recomendações necessárias à implantação de linguagem de indexação interoperável destinada a ser compartilhada em sistemas integrados de bibliotecas universitárias.

## **1.2 Objetivos**

Deste modo, a finalidade desta pesquisa é elencar as características que devem ser preferidas no processo de construção de linguagens de indexação para que esta se constitua em uma linguagem interoperável, permitindo seu intercâmbio e reuso com outros sistemas e bibliotecas universitárias.

Com a realização deste estudo pretende-se alcançar como objetivo geral, oferecer subsídios teóricos e metodológicos à pesquisa sobre a interoperabilidade semântica de linguagens de indexação no contexto teórico da Organização e Representação do Conhecimento.

Para tanto, tem como objetivos específicos:

1. Realizar estudo teórico e metodológico com revisão da literatura nacional e internacional sobre elaboração e manutenção de linguagens de indexação com foco na interoperabilidade semântica;
2. Analisar o processo de elaboração de linguagens de indexação sob o aspecto das normas internacionais focalizando a interoperabilidade semântica;
3. Identificar e analisar projetos de interoperabilidade semântica de linguagens de indexação;
4. Propor recomendações para a construção de linguagens com interoperabilidade semântica.

### **1.3 Justificativa**

O desenvolvimento desta investigação justifica-se diante da importância da construção de linguagens de indexação para os processos de organização do conhecimento em bibliotecas, especialmente diante da difusão de tecnologias que permitem que o processo de construção seja facilitado, bem como, possibilita que o uso destas linguagens seja potencializado, exigindo, desta forma, maior qualidade no planejamento e construção das linguagens de indexação.

Esta pesquisa também se apresenta como continuidade dos estudos iniciados desde o curso de graduação em Biblioteconomia, com a realização de pesquisa de iniciação científica no tema de indexação e política de indexação e prosseguidos no curso de mestrado, em que se abordou a avaliação extrínseca da indexação pela consistência e que demonstrou que a qualidade da indexação em sistemas de bibliotecas universitárias reside em grande parte na qualidade da linguagem de indexação utilizada nestes sistemas.

Esta investigação vincula-se a projeto de pesquisa regular com financiamento concedido pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), cujos objetivos

centrais são “a elaboração de linguagem de indexação para bibliotecas, fornecendo subsídios teóricos e metodológicos para elaboração, implantação e avaliação de linguagens de indexação em ambientes de bibliotecas” (FUJITA, 2015)<sup>3</sup>.

A pesquisa apresenta relevância social ao procurar oferecer subsídios a bibliotecários atuantes na organização e representação do conhecimento em sistemas de bibliotecas universitárias para a construção ou adaptação de linguagens de indexação mais aptas a darem suporte às suas atividades. Visamos com este trabalho de pesquisa proporcionar a este profissional um guia que contribua para a construção ou adaptação de linguagem de indexação dotada de interoperabilidade semântica e apta a ser utilizada de forma compartilhada.

#### 1.4 Sistemática da tese

Após esta introdução, em que buscamos apresentar o tema da pesquisa, delineando seu problema, os objetivos que se pretende alcançar, bem como a hipótese e a tese da pesquisa, o relato da pesquisa seguirá com a exposição das demais seções, como exposto em seguinte.

O segundo capítulo, intitulado “*As linguagens de indexação no contexto da política de indexação: aspectos teóricos*”, apresenta uma contextualização dos tesouros enquanto linguagem de indexação, apresentando o histórico de seu surgimento e desenvolvimento teórico e a noção de tesouro conceitual; a sua estrutura com foco nos processos de construção e manutenção com base na norma ISO 25.964-1 e; sua utilização para indexação e recuperação da informação. Posteriormente, este item abordará uma discussão sobre as características das linguagens de indexação no âmbito da política de indexação, explorando especialmente a questão da escolha entre a adoção de uma linguagem existente ou a elaboração de uma linguagem própria.

O terceiro capítulo, “*Interoperabilidade semântica de linguagens de indexação: panorama das normas e padrões internacionais*”, apresenta e discute o conceito de interoperabilidade semântica, seu objetivo e funções no âmbito das linguagens de indexação com base na norma ISO 25.964-2, apresenta e discute os principais modelos de mapeamentos semântico apresentados pela referida norma como métodos de estabelecimento de interoperabilidade semântica entre linguagens de indexação e trata, também, dos principais formatos e padrões adotados nas bibliotecas para tratamento dos dados de autoridade de assunto.

---

<sup>3</sup> Projeto sob coordenação de Profa. Dra. Mariângela Spotti Lopes Fujita, orientadora desta pesquisa.

Em seguida, o quarto capítulo apresenta as decisões metodológicas que nortearam a realização desta investigação, apresentando o método quadripolar de pesquisa, que articula a realização da pesquisa em quatro polos interconectados, e as técnicas complementares utilizadas, a entrevista semiestruturada e a Análise de Conteúdo de Bardin (2011).

O quinto capítulo, *Apresentação e discussão dos resultados*, abordará os resultados obtidos com a aplicação da técnica de coleta de dados tendo como parâmetro as etapas de análise propostas por Bardin (2011) e agrupadas em categorias de análise.

Por fim, apresentam-se as considerações finais tendo em vista a hipótese apresentada nesta introdução e os objetivos de pesquisa, bem como, os polos de investigação estruturados no método quadripolar.

Em seguida, apresentamos o Quadro 1 que expõe uma representação visual da organização desta pesquisa, onde relacionamos os objetivos específicos e as seções em que serão abordados.

**Quadro 1** - Estrutura sistemática da pesquisa de acordo com seus objetivos

<b>SISTEMATIZAÇÃO DA PESQUISA</b>	
<b>Estrutura</b>	<b>Delimitação</b>
<b>Tema</b>	A interoperabilidade semântica na construção de linguagens de indexação para representação e recuperação do conhecimento em bibliotecas universitárias
<b>Problema</b>	A disseminação de vocabulários com níveis de controle diversos construídos individualmente por bibliotecas universitárias brasileiras como fator de isolamento, por não atenderem às normativas internacionais e assim não permitirem o alcance da interoperabilidade semântica com outros vocabulários.
<b>Proposição</b>	Propõe-se a realizar estudo que proveja recomendações à implantação de linguagem de indexação interoperável destinada a ser compartilhada em sistemas integrados de bibliotecas universitárias.
<b>Objetivo Geral</b>	Oferecer subsídios teóricos e metodológicos à pesquisa sobre a interoperabilidade semântica de linguagens de indexação no contexto teórico da Organização e Representação do Conhecimento.
<b>Capítulo 2</b>	<b>Objetivo específico 1:</b> Realizar estudo teórico e metodológico com revisão da literatura nacional e internacional sobre elaboração e manutenção de linguagens de indexação com foco na interoperabilidade semântica.  <b>Título:</b> <i>As linguagens de indexação: aspectos teóricos e normativos</i>
<b>Capítulo 3</b>	<b>Objetivo específico 2:</b> Analisar o processo de elaboração de linguagens de indexação sob o aspecto das normas internacionais focalizando a interoperabilidade semântica  <b>Título:</b> <i>Interoperabilidade semântica de linguagens de indexação: panorama das normas e padrões internacionais</i>
<b>Capítulo 4</b>	<b>Objetivo específico 3:</b> Identificar e analisar projetos de interoperabilidade semântica de linguagens de indexação  <b>Título:</b> <i>Metodologia</i>
<b>Capítulo 5</b>	<b>Objetivo específico 4:</b> Elaborar recomendações para a construção de linguagens com interoperabilidade semântica  <b>Título:</b> <i>Resultados e Discussões Finais</i>
	<b>Título:</b> <i>Considerações Finais</i>

Fonte: Elaborado pela autora.

## 2 AS LINGUAGENS DE INDEXAÇÃO: ASPECTOS TEÓRICOS E NORMATIVOS

Neste capítulo aborda-se uma contextualização das linguagens de indexação, dando especial ênfase aos tesouros apresentando o histórico de seu surgimento e desenvolvimento teórico, a noção de tesouro conceitual e sua estrutura de modo a perceber como são elaborados, bem como sua utilização para indexação e recuperação da informação. Posteriormente, este item abordará uma discussão sobre as características das linguagens de indexação no âmbito da política de indexação, explorando especialmente a questão da escolha entre a adoção de uma linguagem existente ou a elaboração de uma linguagem própria.

A opção por abordar mais especificamente o tesouro, se dá pela compreensão desta linguagem de indexação como sendo a mais apropriada para o trabalho de Organização do Conhecimento, como um instrumento que permite tanto a representação e recuperação da informação documental como a representação do campo conceitual. Nesse sentido, também Cervantes (2004, p. 35) destaca o tesouro como sendo o mais adequado para as operações informacionais, pois “além de estabelecer relações entre os termos utilizados, denominados de preferidos e não preferidos, este instrumento permite uma visão sistêmica da área - objeto de representação”.

### 2.1 Normas internacionais de construção de linguagens de indexação

No âmbito da Organização do Conhecimento, as linguagens de indexação são instrumentos de controle de vocabulário que objetivam a representação e a recuperação de informação e nesse sentido o papel da normatização internacional é essencial, ao estabelecer diretrizes que orientam para a realização do processo com semelhantes critérios e parâmetros independentemente do contexto nacional ou institucional, de modo que a linguagem “[...] funcione de forma confiável como um guia para a escolha do termo certo para o conceito solicitado” (DEXTRE CLARKE; ZENG, 2012, p. 21, tradução livre).

Dextre Clarke e Zeng (2012) consideram que

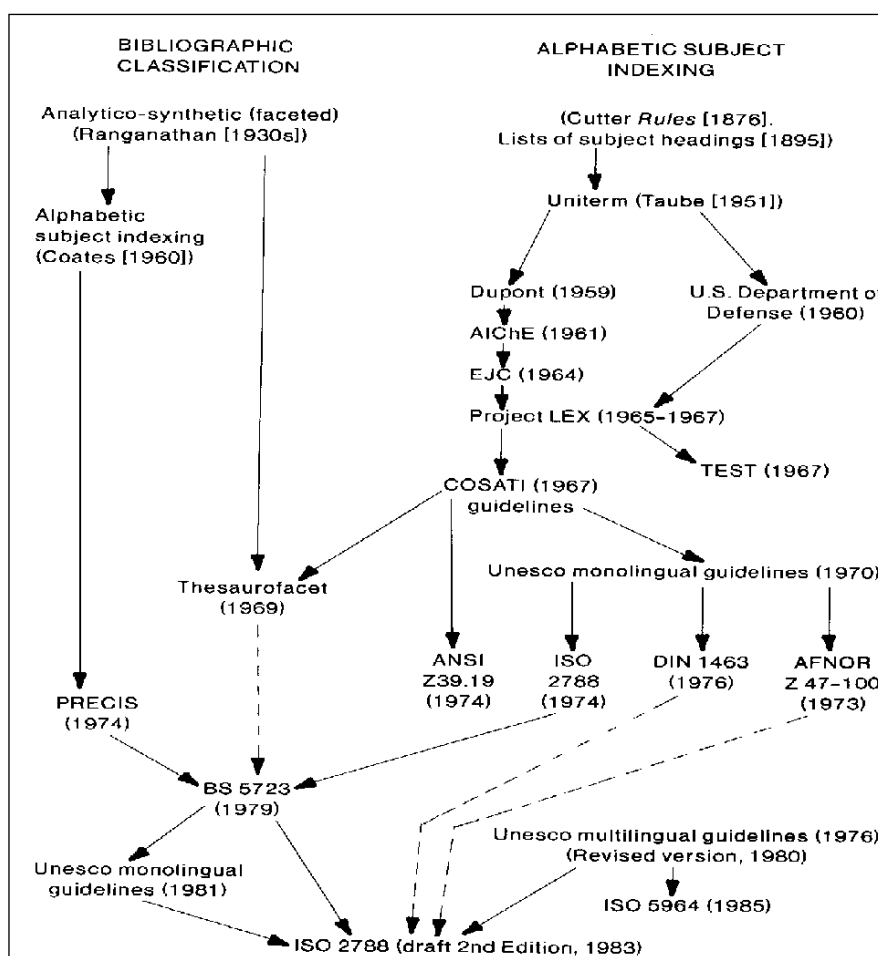
Esta é a base lógica para a abordagem de tesouro: se você pode guiar as pessoas a sempre usar os mesmos termos para os mesmos conceitos e se qualquer termo particular pode se aplicar a apenas um conceito, então os usuários podem pesquisar confiavelmente com palavras, não códigos (DEXTRE CLARKE; ZENG, 2012, p. , tradução livre).



Na evolução das normas de construção, Campos et al. (2006) retomam a análise de Lancaster (1995) que demonstra visualmente a sequência evolutiva de linguagens de indexação e de normas publicadas a este respeito, ressaltando a presença de duas vertentes teóricas, uma de tradição europeia, representativa da teoria sistemática; e uma linha norte-americana, de tradição alfabética, culminando com a norma ISO 2788 (segunda edição, 1986).

A Figura 1 apresenta duas vertentes da construção de tesouros diferentes, porém não totalmente independentes, a vertente europeia apresentada no lado esquerdo, segue a tradição de Ranganathan (década de 1930), baseada na teoria da Análise sintético-facetada, que teve sequência nos estudos do *Classification Research Group*, representado pela Indexação Alfabética de Assunto de Coates (1960) e PRECIS (*PREserved Context Index System*) de Derek Austin (1974)<sup>4</sup>.

**Figura 1** - Evolução das normas e padrões sobre tesouros



Fonte: LANCASTER, 1995, p. 50

<sup>4</sup> AUSTIN, D. The development of PRECIS: a theoretical and technical history. *Journal of Documentation*, London, v. 30, n. 1, p. 47-102, 1974.

Ao lado direito, Lancaster apresenta a vertente norte-americana, decorrente da tradição da indexação alfabética de assuntos (ou catalogação de assuntos) herdeira de Charles Cutter e as Regras para construção de catálogo dicionário (1876)<sup>5</sup> onde Cutter prepara um conjunto de regras para a catalogação nas bibliotecas públicas dos Estados Unidos da América, partindo de constatação das práticas empíricas realizadas nas bibliotecas para padronizá-las. Outro ponto de partida da tradição norte-americana está na Lista de Cabeçalhos de Assunto da Biblioteca do Congresso (*Library of Congress Subject Headings*) que começa a ser elaborada em 1895.

Já em 1951, o próximo marco teórico é o Unitermo de Mortimer Taube, que preconizava a ideia da construção de instrumentos usando apenas termos simples, de uma única palavra, que facilitaria a combinação destes termos para a indexação. De acordo com Lancaster (2004), os unitermos advém da indexação derivada ou indexação por atribuição, em que palavras ou expressões são retiradas do documento de forma literal para serem utilizados na representação de seu conteúdo. O autor destacou que este sistema poderia acarretar problemas como a necessidade de dividir termos como “Oriente Médio”, por exemplo, que se compõe de mais de uma palavra, e a ocorrência de associações falsas ou ambíguas em razão da inexistência de qualquer nível de pré-coordenação.

Posteriormente são desenvolvidos tesouros técnicos voltados a terminologias de áreas especializadas, por empresas e órgãos governamentais, como é o caso do Tesouro do Departamento de Defesa dos EUA (1960). Tais tesouros foram desenvolvidos com base em estudos empíricos e teóricos realizados sob responsabilidade destas organizações. Estes primeiros tesouros e seus desenvolvimentos teóricos e práticos influenciaram a compilação e publicação das diretrizes para construção de tesouros nos EUA, sob responsabilidade do COSATI (Comitê em Informação Científica e Técnica) e foram utilizadas para construir o tesouro TEST (Tesouro de Engenharia e Termos Científicos) publicado em 1967.

A partir daí, sob influência destas diretrizes, neste período percebe-se um esforço de padronização do processo de construção de tesouro, por meio da publicação de diferentes normas como as diretrizes da UNESCO (1970); normas nacionais como a Norma Francesa AFNOR Z 47-100 (1973), Norte-Americana ANSI-NISO Z 39.19 (1974) e Alemã DIN 1463 (1976); e em 1974 a primeira edição da Norma Internacional ISO 2788.

A norma nacional do Reino Unido, BS 5723 é publicada em 1979, influenciada em parte pela Norma Internacional ISO 2788 e pela tradição teórica analítico-sintética.

---

<sup>5</sup> CUTTER, C. A. **Rules for a dictionary catalog**. 3. ed. rewritten. Washington, Govt. printing office, 1891.

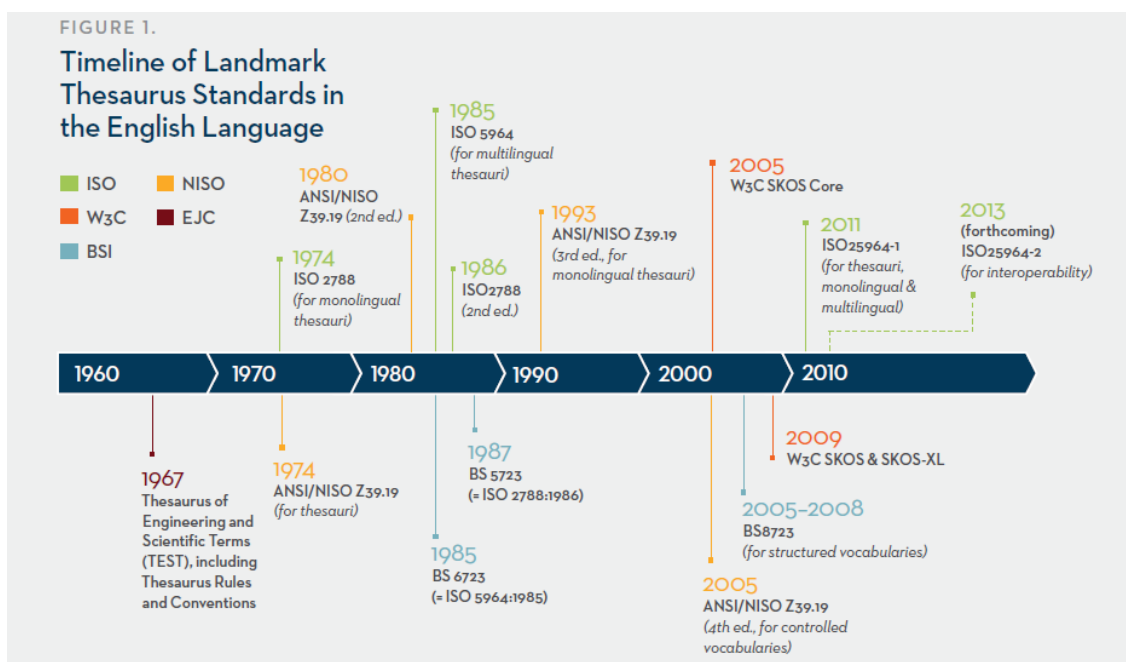
Lancaster (1995) aponta que a evolução destas normas culmina na publicação da segunda edição da Norma Internacional ISO 2788 em 1986, que recebeu influências tanto da tradição europeia quanto da tradição norte-americana.

É importante destacar que neste desenvolvimento histórico são trabalhadas normalmente em normas diferentes a construção de tesouros e vocabulários controlados monolíngues e multilíngues (ISO 5964; BS 6723; UNESCO *Multilingual Guidelines*).

No Brasil, a Norma Internacional ISO 2788 foi traduzida pelo IBICT em 1984, que também traduziu em 1993 as Diretrizes para o estabelecimento de tesouros monolíngues da UNESCO (AUSTIN; DALE, 1993). Outra iniciativa na sistematização dos procedimentos de construção de tesouros no Brasil foi a publicação em 1990 do “Manual de elaboração de tesouros monolíngues” sob coordenação de Gomes (1990).

Posteriormente à Figura 1 apresentada anteriormente, estas normas continuam sendo atualizadas de modo a atender as demandas da comunidade que atua na construção destes instrumentos, recentemente contemplando as questões tecnológicas para a construção de linguagens de indexação apoiada por softwares e a integração destas linguagens entre si e com outras. Tal desenvolvimento é possível observar-se a partir da Figura 2 que trabalha o desenvolvimento da normatização sobre tesouros e linguagens de indexação no idioma inglês.

**Figura 2** - Linha do tempo de marcos em normas e padrões para Tesouros no idioma inglês



Fonte: Dextre Clarke e Zeng (2012, p. 22)

Na Figura 2, são apresentados além das diretrizes TEST, ANSI-NISO Z 39.19, ISO 2788 e ISO 5964, já mencionadas na Figura 1, o desenvolvimento posterior com a publicação em 1986 da ISO 2788, segunda edição da norma Internacional e sua adaptação britânica BS 5723, publicada em 1987. Posteriormente, a ANSI-NISO Z 39.19 recebeu uma terceira edição em 1993. Mas é a partir dos anos 2000 que se percebe maior incremento no desenvolvimento, especialmente com a publicação em 2005 do SKOS Core e em 2009 do formato SKOS e da linguagem SKOS-XL.

Desenvolvimentos estes que estão intrinsecamente conectados à publicação em 2011 da ISO 25.964-1 bem como da ISO 25.964-2, que embora à época da elaboração da Figura 2 por Dextre Clarke e Zeng (2012) não houvesse sido disponibilizada, foi publicada em 2013, tendo sido inclusive objeto de estudo de diversos trabalhos de pesquisa, que vem analisando seus aspectos e sua aplicabilidade, como é o caso de Garcia Marco (2017).

## **2.2 Características dos tesauros e das linguagens de indexação na perspectiva da norma ISO 25.964-1**

Tesauro, em sua acepção mais simples, é um dicionário de ideias afins, um dicionário que em lugar da explicação sobre os significados das palavras, tivesse a preocupação de relacionar as palavras com sentido similar, levando a que o interessado, um autor especialmente, escolha a palavra que melhor expressa o significado pretendido. Essa acepção é a que foi explorada por Peter Roget em seu *“Roget’s Thesaurus of English words and phrases”*, publicado em 1852. Porém esta noção veio a servir de base para os usos futuros do termo tesauro (CURRÁS, 2005).

Foskett já em 1973 destacava que o tesauro ia adiante da simples padronização da representação do assunto, pois ao se apresentar como uma listagem de termos organizados de acordo com as ideias que representam o tesauro permite conhecer a estrutura lógica de um domínio (FOSKETT, 1973).

No âmbito da Ciência da Informação, a palavra tesauro nomeia o instrumento que reúne um conjunto controlado de termos de um determinado domínio, demonstrando as relações semânticas entre estes termos.

O surgimento do tesauro se dá, de acordo com anotação de Currás (2005) em meio ao cenário do pós-guerra na segunda metade da década de 1940 em que começam a ser sentidos os primeiros sintomas do fenômeno chamado de “explosão bibliográfica”. A autora explica que as bibliotecas e centros de documentação especializados encontravam dificuldades de

identificação e localização de informações, tornando-se obsoletos. A flexibilização da representação dos assuntos pelo surgimento em fins da década de 1950 do tesauro com sua estrutura sistemática foi a resposta encontrada para este problema. Neste período os tesouros experimentaram grande evolução teórica e metodológica e passaram a ser utilizados pela Ciência da Informação nos processos de indexação e de recuperação da informação, como o instrumento que reúne um conjunto controlado de termos de um determinado domínio. Dotado de relações sintáticas e semânticas entre termos, o tesauro permite que a linguagem utilizada nos documentos seja traduzida em uma linguagem mais estrita.

Gomes (1990) também atrela essa evolução metodológica permitida pelo tesauro às necessidades informacionais mais especializadas que surgiram a partir dos desenvolvimentos científicos e tecnológicos do pós-guerra.

Era preciso trabalhar com vocabulário mais específico e com uma estrutura mais depurada do que aquela presente nos cabeçalhos de assunto (remissivas e referências cruzadas tipo ver e ver também). Assim, além da especificidade, cuidou-se de melhorar a estrutura e as referências cruzadas (ver também) deram lugar às relações hierárquicas (vertical) e associativas (horizontal) (GOMES, 1990, p. 14).

Também Garcia Marco (2016) reconhece o papel de evidência do tesauro, destacando a relevância do conceito enquanto elemento central desta linguagem,

Eles nasceram como sistemas para superar a diversidade de palavras-chave e unitermos, um sistema inventado para tentar reinar sobre a proliferação de literatura e pesquisas científicas na era das guerras mundiais (que fizeram as classificações bibliográficas parecerem obsoletas), usando a metáfora do dicionário (um sistema alfabético) (GARCIA MARCO, 2016, p. 5).

Fujita e Cervantes (2012) relatam que o termo tesauro foi introduzido na área em 1957, porém encontram-se divergências sobre a sua autoria creditada à Helen Brownson por Vickery (1997)<sup>6</sup> e à Hans Peter Luhn por Lancaster (1995)<sup>7</sup>. Essa adoção se dá em um momento em que houve uma transição da representação do assunto do documento à organização dos conceitos (representados por termos).

Wersig (1971) definiu um tesauro como sendo uma “lista de termos pré-fixados, mas retirados do texto dos documentos, que desdobra os conceitos em unidades simples. Estas

---

<sup>6</sup> VICKERY, B. C. Ontologies. In: *Journal of Information Science*. v. 23, n. 4, p. 277-286, 1997.

<sup>7</sup> LANCASTER, F. W. *El control del vocabulario en la recuperación de información*. Tradução de Alejandro de la Cueva Martín. València: Universitat de València, 1995.

coordenam-se posteriormente para evitar ambiguidades. Entre elas se estabelecem relacionamentos hierárquicos, associativos e de equivalência” (WERSIG, 1971 traduzido por CURRAS, 2005, p. 81).

Para o *United Nations Educational, Scientific and Cultural Organization* (UNISIST) o tesouro pode ser visto e definido segundo duas óticas:

Segundo sua função, é um instrumento de controle terminológico, usado para traduzir de uma linguagem natural dos documentos, indexadores ou usuários a uma “linguagem do sistema” mais restrita. Segundo sua estrutura, são vocabulários controlados e dinâmicos de termos relacionados semântica e genericamente, que cobrem um domínio específico do conhecimento (UNISIST, 1976, p. 5).

O tesouro é compreendido no âmbito da norma ISO 25.964-1 como um “vocabulário controlado e estruturado no qual conceitos são representados por termos, organizados para que as relações entre conceitos sejam feitas explícitas e termos preferidos estejam acompanhados por entradas para sinônimos ou quase-sinônimos” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 12, tradução livre).

De acordo com Lancaster (1995) sua evolução deste momento em diante deu-se com base em duas vertentes teórico-metodológicas, a baseada na análise facetada de Ranganathan e aquela baseada na tradição Norte-americana partindo de Cutter, como explicam Mendes, Reis e Maculan (2015)

[...] Teoria da Análise Facetada, com a análise analítico-sintética, de Ranganathan, teoria que permite organizar e estabelecer representações multidimensionais de um elemento em determinado domínio, característica que foi sendo, ao longo do tempo, assimilada pela estrutura dos tesouros (MENDES; REIS; MACULAN, 2015, p. 63).

O tesouro conceitual é “um tesouro com base em conceitos: seu nome indica que cada termo denota um conceito, ou seja, uma unidade de conhecimento” (CAMPOS; GOMES, 2006, p. 349). Considera-se o tesouro como uma linguagem de indexação cuja prerrogativa é a sua composição que destaca as relações existentes entre termos e conceitos.

Esta concepção de tesouro está baseada na abordagem teórico-metodológica denominada de “abordagem conceitual” que enfatiza a relação entre a teoria do conceito de Dahlberg (1978) e a teoria da classificação, especialmente a herança de Ranganathan, para a construção de tesouros (CAMPOS; GOMES, 2006).

Imbuída pelos princípios da Teoria da Análise facetada, Dahlberg (1978) desenvolve a teoria do Conceito, com a qual os tesouros passam de instrumentos terminológicos para instrumentos conceituais. A autora aponta que os conceitos e as categorias são a base da organização do conhecimento e de sua representação, importantes na criação de tesouros. Como resultado desses fundamentos observamos maior controle nos aspectos linguísticos, na escolha e uso dos descritores para representar os conceitos assim como o estabelecimento das relações semânticas entre os conceitos do tesouro (MENDES; REIS; MACULAN, 2015, p. 63).

Neste processo, a abordagem conceitual de construção do tesouro conceitual adota como unidade de trabalho o conceito e não o assunto - como nos cabeçalhos de assunto, ou o termo - como no tesouro documentário. Para Dahlberg (1978) o conceito é entendido como “unidades de conhecimento, identificadas por meio de enunciados verdadeiros acerca de determinado objeto e representadas por uma forma verbal”.

A abordagem conceitual pressupõe, portanto, o conceito como elemento central da linguagem de indexação, de onde se estabelece a partir do objeto/referente o melhor correspondente para sua designação, na forma do termo, e as relações entre conceitos e termos. Ao abordar esta questão com relação à norma ISO 25964-1 (2011), Dextre Clarke e Zeng (2012) afirmam que por conceito e designação serem inseparáveis, o único modo de um tesouro apresentar sua listagem alfabética de conceitos é com o uso de termos, o que torna inevitável o uso de termos em um tesouro, ou seja, um tesouro irá conter conceitos e termos que os designam.

No intuito de melhor compreender a evolução das características dos tesouros e das compreensões atuais acerca de sua construção e manutenção, é importante delinear a norma internacional vigente para estes instrumentos, a ISO 25.964-1 (2011). Deste modo, em seguida descreve-se a composição da norma ISO 25.964-1 (2011), trazendo os títulos de seus capítulos, os elementos que abordam e um exemplo significativo do conteúdo, quando possível.

A Norma Internacional ISO 25964-1 publicada em 2011, corresponde à atualização e ampliação do conteúdo das normas ISO 2788:1986 (*Documentation -- Guidelines for the establishment and development of monolingual thesauri*) e ISO 5964:1985 (*Documentation -- Guidelines for the establishment and development of multilingual thesauri*), tratando do desenvolvimento e manutenção de tesouros mono e multilíngues. Para completar esse conteúdo a segunda parte da norma, ISO 25964-2, foi publicada em 2013 e aborda a interoperabilidade entre diferentes tesouros e outros vocabulários.

A primeira parte da ISO 25964-1 (2011) é composta por 18 seções direcionadas à construção de tesouros tanto monolíngues como multilíngues, que abordam o tesouro quanto a seu escopo e objetivos; as noções de termo, conceito e conceito complexo; os relacionamentos

entre conceitos; o método de análise facetada para a construção de tesouros; o processo de gestão do tesouro; softwares para construção e gestão de tesouros; modelo de dados para implementação de tesouros por computadores e a integração de tesouros com outras aplicações, bem como os formatos e protocolos para intercâmbio de dados (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011).

As primeiras três seções da norma apresentam elementos técnicos relativos à sua estrutura e vinculação normativa, portanto, a norma apresenta um prefácio e uma introdução e em seguida estão as seções *1 Escopo*, *2 Termos e definições* e *3 Símbolos, abreviaturas e outras convenções*. Em seguida, a seção 4 representa a primeira composta de conteúdo propriamente dito.

A seção 4 *Revisão e objetivos do tesouro* apresenta os objetivos do controle de vocabulário, do uso de tesouros, bem como, os tipos de relacionamentos paradigmáticos e sintagmáticos.

#### **4 Revisão e objetivos do tesouro**

“Estabelecer um termo preferido para representar um conceito particular não é sempre direto porque um conceito pode frequentemente ser expresso de mais de uma forma. Além disso, no discurso comum um termo pode ter mais de um significado. O controle de vocabulário é, portanto, essencial e os tesouros são usados para alcançá-lo” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 16).

A seção 5 *Conceitos e seu escopo em um tesouro* apresenta tipologia e exemplos de conceitos e notas de escopo que podem ser incluídos em tesouros.

#### **5 Conceitos e seu escopo em um tesouro**

“A principal aplicação para um tesouro é na recuperação da informação, onde o objetivo é a busca por conceitos. Como explicado em 4.1 e 4.2, os conceitos são representados por termos. Os conceitos podem variar de simples (por exemplo “gatos”) aos verdadeiramente complexos (por exemplo “discriminação racial entre minorias étnicas”)” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 18).

A seção 6 *Termos do tesouro* apresenta tipos de termos que podem ser incluídos no tesouro e aborda suas formas gramaticais.

#### **6 Termos do Tesouro**

“No discurso comum, um único termo pode ter mais de um significado e o significado mais comum nem sempre é o necessário para representar no tesouro o conceito procurado. [...] Assim como todos os termos do tesouro devem ser expressos o mais inequivocamente possível, é especialmente importante formular o termo preferido de um dado conceito de tal forma que ofereça o escopo pretendido a qualquer usuário” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 21).



A seção 7 *Conceitos complexos* apresenta a conceituação de “conceitos complexos”, instruções para seu uso no tesouro, formas de decomposição de conceitos complexos e instruções sintáticas.

<b>7 Conceitos complexos</b>
“Conceitos complexos são muito frequentemente veiculados por termos compostos, ou seja, termos que podem ser morfológicamente divididos em dois ou mais componentes. Alguns destes são termos multi-palavras e alguns consistem em apenas uma palavra” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 37).

A seção 8 *O relacionamento de equivalência em um contexto monolíngue* aborda o estabelecimento dos relacionamentos de equivalência entre termos de um mesmo idioma do tesouro, indicando quando este relacionamento é estabelecido e as convenções utilizadas para representá-lo.

<b>8 O relacionamento de equivalência em um contexto monolíngue</b>
“Por uma questão de previsibilidade, esses critérios devem ser aplicados consistentemente em todo o tesouro. Se, por exemplo, decidir-se que nomes populares e não nomes científicos devem servir como termos preferidos, essa decisão deve ser aplicada consistentemente, exceto quando um nome popular apropriado não existe. Da mesma forma, convenções de ortografia consistentes devem ser adotadas. As convenções e critérios editoriais devem ser indicados em uma introdução ao tesouro” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 47).

A seção 9 *Equivalência entre linguagens* aborda o estabelecimento dos relacionamentos de equivalência entre termos de linguagens diferentes, podendo aplicar-se à tesouros multilíngues. São apresentados os graus de equivalência entre termos preferidos: equivalência exata, equivalência inexata ou quase-equivalência, equivalência parcial e os casos em que não se pode estabelecer uma equivalência.

<b>9 Equivalência entre linguagens</b>
“Em um tesouro multilíngue, um relacionamento de equivalência inter-linguagens bidirecional deve ser mostrado entre termos preferidos que representam o mesmo conceito em cada um dos idiomas. Como uma opção, cada conceito pode receber uma notação e/ou um identificador único” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 50).

A seção 10 *Relacionamentos entre conceitos* aborda o estabelecimento dos relacionamentos entre conceitos em um tesouro, apresentando sua tipologia e diretrizes de

estabelecimento. São distinguidos os relacionamentos hierárquicos, de gênero-espécie, de todo-parte e instancial e relacionamentos poli-hierárquicos; os relacionamentos associativos; bem como, diretrizes para a customização de relacionamentos hierárquicos e associativos para possibilitar sua subdivisão em tipologias adequadas ao contexto do tesouro, como por exemplo, relacionamentos hierárquicos para siglas e/ou abreviaturas e relacionamentos associativos para causa e efeito.

#### **10 Relacionamentos entre conceitos**

“Além dos relacionamentos de equivalência existem vários outros tipos de relacionamentos que podem ser estabelecidos em um tesouro. Embora os relacionamentos de equivalência em um tesouro monolíngue se aplicam apenas entre *termos* (preferidos e não-preferidos), os relacionamentos introduzidos neste capítulo são estabelecidos entre *conceitos* [...]. Tais relacionamentos são difíceis de apresentar, visto que os conceitos existem apenas na mente e, portanto, a forma convencional de apresentá-los é mostrar cada um destes relacionamentos como aplicados entre os termos preferidos que representam os conceitos em questão.” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 57).

A seção *11 Análise facetada* apresenta diretrizes para a aplicação de análise facetada no arranjo das hierarquias do tesouro. Aborda a utilidade de aplicar a análise facetada, a escolha das facetas, assim como a forma de apresentação, é indicada a possibilidade de incluir os nomes das facetas como termos preferidos tratados como *top terms* (nome da classe mais alta) ou sua inclusão como rótulo de nó.

#### **11 Análise facetada**

“A análise facetada é útil em gerar hierarquias que estão adequadas às regras para relacionamentos hierárquicos, como colocadas em 10.2, porque estes relacionamentos são válidos apenas para conceitos pertencentes a mesma categoria genérica. A escolha de facetas pode variar dependendo da área de assunto, mas nos níveis mais altos é comum usar categorias fundamentais tais como objetos, materiais, agentes, ações, lugares, tempo etc. Onde for útil fazer isso, essas facetas fundamentais podem ser analisadas em sub-facetadas até o nível requerido [...]” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 68).

Na seção *12 Apresentação e leiaute* orientações gerais são feitas para a apresentação de tesouros, bem como orientações específicas para tesouros multilíngues, tais orientações são direcionadas tanto a apresentação de tesouros impressos quanto para visualização de tesouros em formato digital. Aborda as formas de apresentação por registro único, apresentação alfabética, apresentação hierárquica e apresentação classificada.

### **12 Apresentação e leiaute**

Quando apresentado para uso humano, em tela ou em forma impressa, um tesouro pode ser apresentado nas seguintes formas.

- a) Um único registro é a forma mais elementar de apresentação, mostrando o termo preferido ou não-preferido e (opcionalmente) qualquer ou todos os relacionamentos, códigos e notas que se ligam a ele.
- b) Um arranjo alfabético permite acesso aos conceitos a partir das palavras pelas quais são expressos inicialmente pelos usuários.
- c) Uma sequência hierárquica baseada em relacionamentos TG/TE ajudam na expansão ou refinamento do conceito sendo indexado ou buscado.
- d) Uma sequência classificada permite pesquisa em uma área de assunto, direcionando atenção aos conceitos relacionados.
- e) Uma apresentação gráfica mostra termos e seus relacionamentos dispostos pictoricamente.
- f) Uma apresentação permutada ajuda a encontrar palavras incorporadas em termos multi-palavras” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 70).

A seção *13 Gerenciando a construção e manutenção de tesouros* trata de todo o processo administrativo relativo aos tesouros, desde o seu planejamento até os processos de manutenção e atualização que ocorrerão após sua publicação. Portanto, são abordadas as questões que devem ser definidas previamente ao início do trabalho do tesouro, como o contexto de utilização, que envolve os objetivos, os usos planejados e a comunidade de usuários; as características específicas do tesouro, como o formato de apresentação, o formato de disponibilização, entre outras; a disponibilidade de recursos financeiros, humanos e de vocabulário; questões que se referem a elaboração do tesouro em si, como a coleta, compilação e análise dos termos; as opções para construção de tesouros multilíngues; a entrada dos termos e relacionamentos; e as questões relativas à disponibilização do tesouro, tais como a preparação da introdução, a sua disseminação e os procedimentos de manutenção e atualização.

### **13 Gerenciando a construção e manutenção de tesouros**

“Recursos de vocabulários existentes devem ser pesquisados antes da construção iniciar. Qualquer ou todos os seguintes podem provar-se úteis:

- a) Tesouros ou sistemas de classificação existentes com um escopo parcial ou totalmente compatível com o requerido. Em casos extremos, pode ser mais rentável simplesmente adotar um tesouro existente sem mudanças. Mais comumente, porções de um vocabulário existente podem ser tomadas e/ou adaptadas, sujeito a direitos autorais. Trabalhos existentes são fonte de ideias para termos ou estruturas, ou ambos.
- b) Coleções de terminologias ou consultas frequentemente feitas que colegas possam ter montado no curso de seus trabalhos.
- c) Os índices de bases de dados existentes ou outras ferramentas de referências relevantes, que mostrem pontos de entrada úteis para o assunto. Se uma base de dados relevante já existente houver sido indexada com um vocabulário controlado, pode ser possível obter deste um relatório de todos os termos de indexação usados e sua frequência de uso.
- d) O log de transações de um website relevante, que mostre os termos de busca mais empregados pelos usuários.

e) Trabalhos de referência padrão tais como dicionários, terminologias, nomenclaturas e enciclopédias são úteis para verificação de significado de termos e validação de estruturas hierárquicas, assim como um guia para a seleção de termos” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 89-90)
---

A seção 14 *Diretrizes para gestão de softwares de tesouros* apresenta os requisitos fundamentais para as ferramentas de softwares de gestão de tesouros que podem servir de guia quanto aos critérios para a seleção destes softwares. São feitas orientações quanto aos requisitos para indicação de relacionamento entre termos e entre conceitos, notas, códigos, notações, identificadores únicos e rótulos de nó, assim como, para uso de mais de um idioma em um tesouro (multilíngue), importação e exportação de dados e funcionalidades de gestão editorial do(s) tesouro(s).

<b>14 Diretrizes para gestão de softwares de tesouros</b>
---

“Os softwares para a gestão de tesouros devem ser capazes de gerenciar, importar e exportar os dados como descrito na seção 15. Deve ser escolhido com base nos requisitos do projeto em mãos (ver 13.1.5). Embora as características e funções descritas em 14.2 a 14.10 geralmente sejam necessárias, podem ocorrer exceções. Além disso, recursos adicionais são frequentemente necessários. Por exemplo, pode ser vantajoso se o software é capaz de manusear simultaneamente mais de um tesouro” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 98).
--

A seção 15 *Modelo de dados* apresenta um modelo de dados em UML (*Unified Modeling Language*) para intercâmbio de tesouros. O modelo orienta para uma estrutura lógica que visa atender às características dos tesouros desenhadas nas seções anteriores desta parte da norma. Nesta seção da norma o modelo é apresentado em formato de diagrama e de tabela para maior especificação de suas convenções.

<b>15 Modelo de dados</b>
---------------------------

“Para o intercâmbio entre tesouros ser bem sucedido em uma ampla gama de situações, é importante ter um formato hospitaleiro tanto a tesouros simples bem como sofisticados. O modelo apresentado aqui acomoda a completa gama de opções descritas nas seções 5 a 12. Alguns elementos deste modelo são mostrados como opcionais e não necessitam ser usadas quando intercambiando tesouros que não contém as características opcionais.” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 103)
--

A seção 16 *Integração de tesouros com aplicações* descreve características que irão permitir a interoperabilidade dos tesouros com sistemas em rede que gerenciam tesouros e/ou os sistemas de indexação e pesquisas em que estes serão utilizados, bem como, características que devem ser preferidas nestes sistemas de indexação e pesquisa.

### **16 Integração de tesouros com aplicações**

“Para aplicações de pesquisa e indexação, exceto quando essas estão totalmente automatizadas sem preparação para interação de usuários, facilidades para pesquisa e consulta ao tesouro são fundamentais. Estas devem incluir as seguintes:

- a) A habilidade de pesquisar por termos contendo qualquer palavra, combinação de palavras ou sequência de caracteres.
- b) A habilidade de pesquisar por termos e conceitos usando qualquer notação e/ou identificadores presentes.
- c) A habilidade de escolher ou trocar o idioma apresentado de um tesouro multilíngue, ou limitar a busca e apresentação para uma única linguagem especificada.
- d) A habilidade de apresentar os equivalentes (preferidos ou não preferidos) para um dado termo, incluindo equivalentes interidiomáticos se o tesouro é multilíngue.
- e) A habilidade de passar de um registro para um conceito para aquele de qualquer um de seus conceitos genéricos, específicos ou relacionados, geralmente alcançado ao configurar links hipertextuais. Essa facilidade pode remover a necessidade de uma notação para vincular as apresentações hierárquicas ou alfabéticas e classificadas.
- f) A habilidade para expandir e contrair níveis quando visualizando uma apresentação classificada ou hierárquica. Para conseguir uma visão geral inicial da estrutura deve ser possível iniciar com uma visão do nível maior, então expandir progressivamente e seletivamente para baixo. Deve também ser possível expandir ou contrair totalmente em um movimento.
- g) A habilidade de passar de um estilo de apresentação para outro, por exemplo, da apresentação classificada para a alfabética. Preferencialmente deve ser possível visualizá-los simultaneamente.” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 117).

A seção *17 Formatos de intercâmbio* apresenta os formatos conhecidos de representação de dados de autoridade. São apresentados brevemente o formato MARC21, para dados de autoridade e de classificação, MADS (Esquema de Descrição de Autoridades em Metadados), SKOS, Zthes e DD 8723-5. Porém, esta sessão da norma reforça que nenhum dos formatos existentes está adequado plenamente para as características que os tesouros construídos com o modelo de dados exposto na sessão 15 irá possuir e dessa forma, a norma recomenda, por um lado, o uso do formato XML que acompanha o modelo de dados e por outro, esquemas de transformação de dados para conversão dos dados como forma de minimizar o problema do uso de múltiplos formatos.

### **17 Formatos de intercâmbio**

“A interoperabilidade entre diferentes aplicações computacionais requer o uso de um formato comum [...]. Como nenhum dos formatos existentes é adequado para conter todas as características de um tesouro usando o modelo na Seção 15, o formato XML descrito no Anexo B é recomendado.” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 118-119).

A seção *18 Protocolos* apresenta os protocolos de transmissão de dados, que são ferramentas essenciais para que os tesouros possam funcionar em redes digitais comunicando-se com outras aplicações, de modo que se torna uma ferramenta de apoio a interoperabilidade

semântica. Os protocolos são divididos pela norma em protocolos de uso geral e protocolos específicos para tesouros, explicando que embora os protocolos específicos para tesouros estejam mais preparados para explorar as características estruturais do tesouro, os protocolos de uso geral podem facilitar a integração com outras aplicações existentes bem como é mais provável encontrar desenvolvedores familiarizados com eles. Dessa forma, a escolha do protocolo deve embasar-se nas necessidades do projeto e nos recursos disponíveis. São apresentados dois exemplos de protocolos específicos para tesouros: SWAD-E SKOS API, é uma API (Interface de programação de aplicações) para serviços Web planejada para permitir acesso a tesouros e outras linguagens SKOS via Web e o *ADL Thesaurus Protocol (Alexandria Digital Library)* é um protocolo para acessar tesouros monolíngues, adaptado para uso na Web.

#### 18 Protocolos

“Alguns protocolos existentes foram planejados especificamente para tesouros enquanto outros são protocolos de propósito geral que estão sendo usados (ou adaptados) para trabalhar com tesouros. O primeiro tipo permite explorações mais poderosas da estrutura e relacionamentos do tesouro, enquanto os instrumentos de propósito geral podem facilitar a integração com outras funcionalidades ou conjuntos de ferramentas existentes. Pode também ser possível aplicar uma interface ou perfil de tesouro a um protocolo de propósito geral. Em alguns casos, o conjunto limitado de funções oferecido por um protocolo de tesouro específico pode oferecer vantagens ao abrigar detalhes da representação ou arquitetura subjacente. Por outro lado, às vezes a grande flexibilidade de um protocolo de propósito geral pode ser explorada por desenvolvedores familiarizados a ele” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p. 120).

### 2.3 Escolha da Linguagem de indexação: construção de uma linguagem própria ou adoção de uma linguagem existente

A política de indexação comporta orientações para a tomada de decisões no planejamento de sistemas de recuperação da informação, de forma a garantir que qualquer informação seja provida ao usuário no momento preciso, da maneira mais eficiente e econômica possível (CARNEIRO, 1985). Para a adoção de uma política de indexação são definidas as decisões do sistema quanto aos elementos componentes da política, que na visão de Carneiro (1985) são a cobertura de assuntos, seleção e aquisição de documentos-fonte, o processo de indexação (composto por nível de exaustividade, nível de especificidade, escolha da linguagem e capacidade de revocação e precisão do sistema), estratégia de busca, tempo de resposta do sistema, formato de saída dos dados e avaliação do sistema.

Na definição da política de indexação, a escolha da linguagem destaca-se como um de seus principais elementos, pois a qualidade da linguagem afeta o desempenho do sistema de

recuperação de informação, visto que ela irá nortear o indexador na seleção do termo que melhor representa o conteúdo do documento, bem como, guiará o usuário na formulação de estratégia de busca (CARNEIRO, 1985).

Ao realizar a escolha da linguagem diversos fatores devem ser considerados, o primeiro deles deverá ser entre o uso de uma linguagem livre, sem controle de vocabulário em que serão empregadas as próprias palavras do autor para descrição do assunto do documento, ou uma linguagem de indexação, que compatibiliza as linguagens do autor, do indexador e do usuário. Carneiro (1985) sugeria que o uso de uma linguagem livre em bibliotecas pequenas com coleção especializada e sistema automatizado poderia ser vantajoso ao reduzir os custos da indexação, por torná-la um procedimento mais rápido. Porém a autora também alertava para as dificuldades que poderiam surgir no momento da busca, por ser necessário lidar com as ambiguidades inerentes à linguagem natural. Com a evolução das tecnologias empregadas na automatização dos processos informacionais, os sistemas de recuperação da informação em texto completo que podem funcionar com indexação automática surgem como outra opção que pode ser considerada. Entretanto, ainda é preciso ponderar a questão do ruído e do silêncio na recuperação da informação, causado pela falta de padronização.

Já a linguagem de indexação oferece a vantagem do controle de vocabulário, em que são empregadas técnicas para eliminar a sinonímia e as diferenças de grafia (singular e plural, gênero etc.), o uso de remissivas para direcionar o usuário ao sinônimo aceito na linguagem, além de possuir uma rede de relacionamentos semânticos que objetiva construir uma representação das áreas temáticas, servindo como orientação a indexadores e usuários. Porém o uso de linguagem controlada pressupõe investimentos institucionais na construção/aquisição e manutenção da linguagem, bem como, em treinamentos para indexadores e usuários. Além de haver o risco da desatualização da linguagem, e com isso, sua inadequação e/ou inconsistência causarem problemas de representação da informação, ao não ser uma ferramenta que auxilie verdadeiramente o indexador, e problemas de recuperação de informação, tais como o ruído e o silêncio, levando a redução da confiança dos usuários no sistema, conforme relatado por Boccato e Fujita (2010).

Eleito o uso de linguagem controlada para indexação, o sistema deverá considerar entre as tipologias das linguagens de indexação, sendo comum considerar entre linguagens pré e pós-coordenadas. Nas linguagens pré-coordenadas os conceitos que a compõem são combinados no momento representação do documento pelo indexador, que decide qual a ordem de citação dos termos a partir de regras determinadas pelo sistema. Já nas linguagens pós-coordenadas, o

indexador representa o documento atribuindo separadamente cada conceito apropriado e a combinação dos conceitos é feita posteriormente pelo usuário, na estratégia de busca.

As linguagens pré-coordenadas, tais como as listas de cabeçalhos de assunto, oferecem a vantagem de permitir a representação de assuntos bastante específicos e conceitos complexos por meio da construção de cadeias de assunto, mas em catálogos online, em que a busca é feita por formulação de uma estratégia pelo próprio usuário, essa complexidade se coloca como um obstáculo já que requer do usuário que conheça a forma exata como o cabeçalho foi elaborado pelo indexador.

Já as linguagens pós-coordenadas presentes nos sistemas de informação em âmbito digital, como os tesouros, atribuem aos usuários certa liberdade na elaboração de estratégias de busca, e com isso, maior independência na busca por informação, em geral garantindo maior relevância e exaustividade na recuperação da informação, porém podendo levar a falsas combinações (MARTINS, 2013, p. 51-53).

Outra questão que precisa ser ponderada na escolha da linguagem de indexação em um sistema diz respeito à opção por uma linguagem existente ou a construção de uma linguagem própria. Definidas as características desejáveis na linguagem de indexação, é preciso considerar que a construção de linguagens de indexação deve ser feita “[...] com base na terminologia de uma área, com categorias e subcategorias delimitadas, a partir de relações lógico-semânticas de termos claros e bem definidos, possibilitando uma representação adequada para a recuperação da informação” (BOCCATO; FUJITA, 2010, p. 26).

Construir uma linguagem requer, portanto, um grande investimento em planejamento por parte da instituição, de modo a conhecer principalmente: as necessidades específicas dos usuários; as características das áreas de assunto abrangidas pela coleção que será necessário indexar; e as bases teóricas e metodológicas das linguagens de indexação, visto que a construção de linguagens de indexação é bastante complexa e requer metodologias bem esclarecidas sobre sua execução, “[...] sobretudo no que se refere à construção da estrutura de categorias que define e delimita o campo semântico” (FUJITA, 2011, p. 37).

Desta forma, convém considerar a existência de linguagens que atendam às necessidades de cobertura temática desejada, que tenha atualização periódica ou frequente, que atendam as características estruturais já abordadas aqui, tais como os critérios de controle de vocabulário, o momento da coordenação dos termos, a rede de relacionamentos hierárquicos e semânticos, bem como o formato de acesso e o atendimento às normas internacionais de construção. A opção pela adoção de uma linguagem existente pode significar economia de recursos financeiros e humanos e tem embasamento na filosofia de cooperação responsável por



grande parte dos avanços na área de organização do conhecimento, vindo também ao encontro das recentes iniciativas de dados abertos e interligados.

Assim como a questão da adoção de um instrumento de controle de vocabulário e sobre qual linguagem de indexação específica adotar, a política de indexação também comporta a decisão sobre os aspectos de exaustividade e especificidade desejados na indexação, que terão efeitos na recuperação da informação no catálogo, refletindo em sua qualidade.

Em estudos anteriores (PIOVEZAN, 2015) realizamos análise e sistematização conceitual destes elementos abordando sua relação entre si e seus efeitos na consistência da indexação e na precisão e exaustividade da recuperação da informação.

Porém, para que os níveis de especificidade e exaustividade pretendidos, definidos na política de indexação, e os níveis de exaustividade e especificidade reais alcançados no catálogo, conforme explicado por Soergel (1994), estejam em sintonia é preciso considerar tais elementos não apenas no contexto da indexação, mas também no contexto da estrutura das linguagens de indexação.

Como um instrumento de auxílio tanto na representação do conteúdo de documentos por parte do indexador assim como na escolha de uma estratégia de busca para recuperação de informação por parte do usuário, a linguagem de indexação exerce função complementar e integradora no sistema, ao perpassar suas funções desde a incorporação no sistema até o acesso à informação. Essa afirmação também pressupõe que a recuperação da informação será inequívoca se realizada com uma linguagem de indexação apropriada, comum a usuário e indexador.

O princípio básico para que uma linguagem de indexação atue de modo inequívoco é o da univocidade de significação, ou seja, que uma mesma palavra usada na representação possua o mesmo significado no documento; para o indexador, que alimenta o sistema; e para o usuário que irá elaborar sua estratégia de busca.

Do mesmo modo, a linguagem também necessita de coesão e consistência conceitual, ser dotada de relacionamentos seguros entre os termos e deverá possuir satisfatória exaustividade e especificidade em suas áreas de assunto, de modo que oriente indexador e usuário a escolher apropriadamente entre os termos que a compõem.

A exaustividade e a especificidade têm reflexos observáveis na recuperação da informação, uma indexação com alta exaustividade tende a aumentar a quantidade de resultados recuperados na busca, oferecendo, portanto, uma recuperação também exaustiva, porém com menor grau de precisão, ou seja, de menor acerto na apresentação de resultados desejados pelo usuário.

Em contrapartida, uma indexação com alta especificidade tende a anular resultados considerados inúteis ou irrelevantes na busca, ou seja, tende a oferecer resultados mais precisos ao usuário, porém reduzindo a quantidade de documentos recuperados.

A especificidade e a exaustividade são, por estas razões, consideradas elementos chave na definição e elaboração da política de indexação, pois ao influenciarem a recuperação da informação, também influenciam a confiança que o sistema inspira no usuário.

Estas reflexões nos levam a ponderar estes elementos e a revisar as considerações teóricas oferecidas sobre eles na literatura, com atenção à construção de linguagens de indexação, visto que, a indexação exaustiva ou específica poderá ser definida na política de indexação, mas também dependerá da existência de termos apropriados na linguagem, bem como da existência de estrutura hierárquica e semântica em que este termo apropriado possa ser corretamente identificado.

Fujita (2012, p. 25) coloca que a exaustividade possui dois pontos de vista relatados na literatura, são eles:

- a) o aspecto quantitativo no qual a exaustividade se refere ao número de termos atribuídos pelo indexador (ABNT, 1992; LANCASTER, 2004) e
- b) o aspecto de extensão da representação do conteúdo coberto pelos termos atribuídos pelo indexador (SOERGEL, 1994; GIL LEIVA, 2008).

Para atender aos dois pontos de vista relatados, a linguagem de indexação precisa ser extensiva na inclusão de termos e remissivas, que ofereçam ao indexador a possibilidade de expandir a representação do documento com o maior número de conceitos identificados no conteúdo do documento; assim como possuir termos com a capacidade de cobrir todo o conteúdo compreendido pelo documento.

Percebe-se que o primeiro ponto de vista de Fujita (2012) dirige-se para a realização da indexação em profundidade, em que se representarão os subtemas tratados no documento, o que requer da linguagem que seja exaustiva no sentido de completude, ou seja, de possuir todos os termos necessários para representar fidedignamente os conteúdos dos documentos a indexar.

A essa noção assemelham-se as definições apresentadas por Van Slype (1991, p. 123) que coloca a exaustividade como “a qualidade na eleição dos conceitos realmente significativos, ou seja, que contém informação pertinente aos usuários” (VAN SLYPE, 1991, p. 123, tradução livre).

A definição proposta por Wellisch (1995, p. 175, tradução livre) também se aproxima desta noção, afirmando que a exaustividade “se refere à extensão com a qual conceitos e tópicos são feitos recuperáveis por meio de termos de indexação”. A concepção do autor também

relembra que os conceitos que não puderem ser representados por termos de indexação serão omitidos ao usuário em seu processo de busca e os documentos provavelmente não serão recuperados em questões para as quais seria útil.

Lancaster (2004) aponta para a questão quantitativa e assevera o efeito da exaustividade na recuperação da informação, segundo o autor a exaustividade se refere ao “[...] emprego de termos em número suficiente para abranger o conteúdo temático principal do documento. Quanto mais termos forem utilizados para indexar um documento, mais acessível ele se tornará e, provavelmente, mais vezes será recuperado” (LANCASTER, 2004, p. 27).

Olson e Boll (2001) compreendem a exaustividade como “o número de conceitos representados no registro bibliográfico ou a abrangência do assunto coberto” (2001, p. 91, tradução livre), definição que corrobora a noção exposta por Fujita (2012) apontando para os dois aspectos da exaustividade.

O segundo ponto de vista apresentado por Fujita (2012) aponta para a realização da indexação seletiva, em que se opta pela indicação do tema global do documento, que em geral poderá ser expresso por um termo genérico na estrutura hierárquica.

Além de Fujita (2012) e Olson e Boll (2001), Soergel (1994, p. 591) também orienta a compreensão deste aspecto da exaustividade. De acordo com o autor “Uma entidade sendo indexada é relevante, em diferentes graus, para uma série de conceitos. A exaustividade da indexação é a extensão com a qual estes conceitos são cobertos pelos descritores atribuídos à entidade”.

Procuramos sintetizar as definições encontradas na literatura para a exaustividade da linguagem de indexação para uma melhor visualização dos diferentes pontos de vista sobre tal conceito pelos teóricos de Organização do Conhecimento. O quadro 2 demonstra esta síntese. Tendo sido elaborado com objetivo de uma análise conceitual de exaustividade da indexação para estudos de avaliação da indexação, considerou-se que seria adequado utilizar sua base explicativa, porém adaptando-o aos atuais objetivos referentes a exaustividade nas linguagens de indexação.

**Quadro 2 - Análise conceitual – exaustividade**

<b>Exaustividade</b>	
<b>LANCASTER</b>	“emprego de termos em número suficiente para abranger o conteúdo temático principal do documento. Quanto mais termos forem utilizados para indexar um documento, mais acessível ele se tornará e, provavelmente, mais vezes será recuperado” (LANCASTER, 2004, p. 27)
<b>OLSON; BOLL</b>	“é o número de conceitos representados no registro bibliográfico ou a abrangência do assunto coberto” (OLSON; BOLL, 2001, p. 91)
<b>WELLISCH</b>	“extensão com a qual conceitos e tópicos são feitos recuperáveis por meio de termos de indexação” (1995, p. 175)
<b>SOERGEL</b>	“Uma entidade sendo indexada é relevante, em diferentes graus, para uma série de conceitos. A exaustividade da indexação é a extensão com a qual estes conceitos são cobertos pelos descritores atribuídos à entidade” (SOERGEL, 1994, p. 591)
<b>CARNEIRO</b>	“capacidade do sistema em assegurar a recuperação de um número desejável de documentos relevantes” (CARNEIRO, 1985, p. 234)
<b>SOERGEL</b>	“a fração de todos os itens relevantes corretamente recuperados [...] pode ser vista como a probabilidade de um item ser recuperado, dado que é relevante” (SOERGEL, 1994, p. 590)
<b>LANCASTER</b>	“a capacidade do sistema para recuperar documentos relevantes da base de dados, em resposta a uma petição de informação” (1995, p. 151)
<b>GIL URDICIAIN</b>	“a exaustividade mede a capacidade do sistema para recuperar documentos úteis” (1997, p. 277)

**Fonte:** Adaptado de Piovezan (2015).

Para Olson e Boll (2001) a especificidade se refere ao “nível de detalhamento da terminologia em um vocabulário em termos hierárquicos” (OLSON; BOLL, 2001, p. 95, tradução livre). A especificidade, para os autores, se divide em três aspectos diferentes: a especificidade e co-extensividade da linguagem, a especificidade de aplicação e a especificidade do termo no contexto de indexação.

O primeiro aspecto, a especificidade e co-extensividade da linguagem “determina o quão detalhada pode ser a representação em termos hierárquicos” (OLSON; BOLL, 2001, p. 96, tradução livre) e diz respeito à correspondência semântica entre o termo que está disponibilizado na linguagem e o conceito presente no documento. Os autores sugerem que uma forma de aumentar o nível de especificidade, no que se refere a esse aspecto, é a inclusão de subdivisões a um termo.

Pode-se identificar que as definições propostas por Foskett (1973), Lancaster (1995), Wellisch (1995), Gil Urdiciain (2004) e Rubi (2012) situam-se neste aspecto da especificidade, que requer da linguagem riqueza de detalhamento das suas hierarquias, especificando os conceitos por meio do arranjo em conceitos superordenados e subordinados.

Foskett (1973, p. 12) definiu a especificidade como “a extensão em que o sistema nos permite ser precisos ao especificarmos um documento que estejamos processando”, ressaltando a relação existente entre especificidade e precisão.

Lancaster afirmou que “a indexação e a recuperação podem ser tão específicas quanto permita a linguagem do sistema” (LANCASTER, 1995, p. 159, tradução livre). Para o autor a especificidade é a característica mais importante da linguagem, por determinar a precisão da recuperação e por controlar o tamanho das categorias da linguagem.

A especificidade para Wellisch (1995) “se refere à extensão com a qual um conceito ou tópico em um documento é identificado por um termo preciso na hierarquia de suas relações de gênero-espécie” (WELLISCH, 1995, p. 439, tradução livre). O autor ainda explica que “enquanto a exaustividade é essencialmente uma questão de política de indexação, a especificidade, próximo à terminologia do texto em si, depende da linguagem de indexação utilizada” (WELLISCH, 1995, p. 439, tradução livre).

Gil Urdiciain (2004, p. 29, tradução livre) coloca que a especificidade “se refere à importância de utilizar os conceitos sob seu nome mais específico e não os englobar em um mais genérico”.

Rubi (2012, p. 112) também corrobora este aspecto e define a especificidade da indexação “[...] no que diz respeito à co-extensividade do vocabulário, ou seja, a linguagem utilizada deve permitir que os termos selecionados para indexação sejam traduzidos de maneira efetivamente específica”.

O segundo aspecto proposto por Olson e Boll (2001) é o da especificidade de aplicação, que como o próprio nome sugere, é o “nível de detalhamento com o qual o vocabulário é aplicado” (OLSON; BOLL, 2001, p. 95, tradução livre), cuja prática comum derivada do princípio de Cutter (1891)<sup>8</sup> indica a utilização do termo mais específico existente na linguagem para representar um dado conceito.

Nessa visão de especificidade são identificadas as definições propostas por Van Slype (1991) e Soergel (1994). Nesse aspecto se argumenta sobre a necessidade de que indexadores conheçam a estrutura da linguagem de indexação de modo realizar a tradução dos conceitos identificados no documento com respeito a seu nível hierárquico.

Van Slype (1991, p. 123, tradução livre) afirma que a especificidade “[...] mede a qualidade na eleição dos descritores que correspondem efetivamente aos conceitos incluídos no documento”, referindo-se à exatidão na tradução do conceito para descritores com relação ao nível hierárquico e sugerindo o autor que o conceito seja traduzido para um termo no mesmo nível hierárquico ou no nível imediatamente superior na hierarquia.

---

<sup>8</sup> CUTTER, C. A. Rules for a dictionary catalog. 3. ed. rewritten. Washington, Govt. printing office, 1891.

Soergel (1994, p. 593, tradução livre) define a especificidade como “o nível genérico em que os conceitos atribuídos à entidade são expressos”. Para o autor pode haver diferenças em um sistema entre a especificidade pretendida, aquela definida como padrão pela política de indexação da instituição, e a especificidade real do catálogo em razão de erros de indexação ou do uso de uma linguagem inapropriada.

Van Slype (1991) também se refere à especificidade horizontal, que seria a tradução de conceitos compostos para descritores pré-coordenados em preferência à associação de descritores simples. Essa concepção associa-se a noção de especificidade do termo apontada por Olson e Boll (2001).

O terceiro aspecto, a especificidade do termo é dependente do contexto de indexação e, por isso, pode variar de catálogo para catálogo em razão da cobertura temática do acervo, especialmente sua especialização, sendo que um termo considerado geral em um catálogo pode ser um termo específico dentro da hierarquia de outro catálogo. Olson e Boll (2001) definem esse aspecto da especificidade como “a especificidade relativa de um vocabulário em um dado catálogo ou base de dados” (OLSON; BOLL, 2001, p. 98, tradução livre).

O Quadro 3 expõe uma apresentação visual das definições encontradas na literatura para o conceito de especificidade, de modo a obter uma melhor compreensão dos três aspectos identificados quanto a este conceito na área de organização do conhecimento. Tendo sido elaborado com objetivo de uma análise conceitual de especificidade e precisão da indexação para estudos de avaliação da indexação, considerou-se que seria adequado utilizar sua base explicativa, porém adaptando-o aos atuais objetivos referentes a exaustividade nas linguagens de indexação.

Quadro 3 - Análise conceitual – especificidade

Especificidade		
Especificidade e co-extensividade da linguagem	<b>OLSON; BOLL</b>	“é o nível de detalhamento da terminologia em um vocabulário em termos hierárquicos” (OLSON; BOLL, 2001, p. 95)
	<b>FOSKETT</b>	“a extensão em que o sistema nos permite ser precisos ao especificarmos um documento que estejamos processando” (1973, p. 12)
	<b>LANCASTER</b>	“a indexação e a recuperação podem ser tão específicas quanto permita a linguagem do sistema” (LANCASTER, 1995, p. 159, tradução livre).
	<b>WELLISCH</b>	“se refere à extensão com a qual um conceito ou tópico em um documento é identificado por um termo preciso na hierarquia de suas relações de gênero-espécie” (WELLISCH, 1995, p. 439)
	<b>GIL URDICIAIN</b>	“se refere à importância de utilizar os conceitos sob seu nome mais específico e não os englobar em um mais genérico” Gil Urdiciain (2004, p. 29)
	<b>RUBI</b>	“[...] no que diz respeito à co-extensividade do vocabulário, ou seja, a linguagem utilizada deve permitir que os termos selecionados para indexação sejam traduzidos de maneira efetivamente específica” Rubi (2012, p. 112)
Especificidade de aplicação	<b>OLSON; BOLL</b>	“nível de detalhamento com o qual o vocabulário é aplicado” (OLSON; BOLL, 2001, p. 95)
	<b>SOERGEL</b>	“o nível genérico em que os conceitos atribuídos à entidade são expressos” (1994, p.593)
	<b>RUBI</b>	“nível de abrangência em que o sistema permite especificar os conceitos identificados documento” (RUBI, 2008, p. 45)
	<b>VAN SLYPE</b>	“a especificidade mede a qualidade na eleição dos descritores que correspondem efetivamente aos conceitos incluídos no documento” (VAN SLYPE, 1991, p. 123)
	<b>MARTINS</b>	“Exatidão com que um determinado conceito é representado por um termo de indexação” (MARTINS, 2014, p. 158).
Especificidade do termo	<b>OLSON; BOLL</b>	“a especificidade relativa de um vocabulário em um dado catálogo ou base de dados” (OLSON; BOLL, 2001, p. 98)

Fonte: Adaptado de Piovezan (2015).

Diante das considerações expostas, compreende-se que são ainda necessários aprofundamentos sobre a questão da escolha da linguagem de indexação, especialmente no que se refere à opção entre construir uma linguagem própria e adotar uma linguagem existente. Este tema merece nossa atenção dado o fato de que temos consciência da complexidade da empreitada que significa construir uma linguagem de indexação, mesmo uma linguagem pequena destinada a bibliotecas, assim como outras unidades de informação, de pequeno porte.

E o que pode ser dito empiricamente é que não é acanhado o número de bibliotecas que opta por adotar como padrão o uso de uma linguagem de indexação existente, construída por outra instituição, como pode ser observado no panorama apresentado por Cruz, Santos e Fujita (2016). Porém também é crescente a opção por desenvolver uma linguagem própria, como abordamos anteriormente. Não se pode dizer que entre estas haja uma opção mais correta, desde que observados os critérios de adequação necessários a cada unidade de informação, ao decidir-se pelo uso de uma linguagem existente; bem como sejam atendidas as normas vigentes para a construção de linguagens de indexação, no caso da opção pela elaboração de uma linguagem própria.

Considerando estas questões, especialmente sobre a existência de diferentes linguagens de indexação díspares atualmente nos cenários nacional e internacional, é imprescindível considerar a hipótese de lidar com a disponibilidade de diferentes linguagens e a possibilidade de integrá-las.



### 3 INTEROPERABILIDADE SEMÂNTICA DE LINGUAGENS DE INDEXAÇÃO: PANORAMA DAS NORMAS E PADRÕES INTERNACIONAIS

Neste capítulo será apresentado e discutido o conceito de interoperabilidade semântica e seus objetivos no âmbito das linguagens de indexação, será abordado também o mapeamento semântico como método de estabelecimento de interoperabilidade semântica, apresentando e discutindo três modelos de mapeamento. Este capítulo tem como embasamento principal a norma internacional de construção de tesouros ISO 25.964-2, cuja especificidade centra-se no estabelecimento de interoperabilidade entre tesouros e outras linguagens de indexação e, dessa forma expõe também um panorama desta norma.

Interoperabilidade é uma denominação atual para uma função muito antiga perseguida por profissionais da informação para atender a uma demanda, tão antiga quanto, por identificar e ter acesso a informação documental que está dispersa em diversos ambientes. Se no passado tratava-se de reunir dados sobre documentos custodiados por bibliotecas da antiguidade e construir catálogos que seriam copiados a mão para dar a conhecer a estudiosos a localização onde a fonte desejada deveria ser procurada, no cenário contemporâneo esta informação é digital e está disposta em incontáveis bases de dados, repositórios, bibliotecas digitais e catálogos, que tentam reunir, gerenciar e dar acesso a um número cada vez mais crescente de documentos para usuários que se encontram igualmente dispersos.

Os pesquisadores e usuários de bibliotecas atualmente, acostumados à velocidade e mobilidade da web, têm demandas e expectativas quanto aos serviços de informação que desafiam os catálogos das bibliotecas. Os catálogos públicos de acesso online (OPACS – *Online Public Access Catalogs*) são uma realidade bastante difundida e, se em seus dias iniciais permitiam que os usuários pudessem descobrir de forma remota o que a biblioteca dispunha sobre um determinado tema; atualmente, com a aquisição de bases de dados, a criação de repositórios institucionais e temáticos, estes catálogos permitem aos usuários a partir de seus próprios dispositivos consultar diversas coleções e acessá-las, localizando documentos que podem ser lidos online, transferidos por meio de *download* ou mesmo impressos.

Esta possibilidade de acesso e uso representa um diferencial nas bibliotecas universitárias, porém também tem sido ofertada por mecanismos de buscas generalistas, como o sistema *Google*, e nesse contexto oferecer ferramentas de organização e de recuperação do conhecimento de difícil acesso, mesmo que disponham de coleções especializadas e com qualidade assegurada, pode significar que estas ferramentas não serão utilizadas. O mesmo se

pode dizer quanto ao uso de catálogos e bases de dados que indexam o conteúdo de documentos com diferentes vocabulários especializados.

Como já expressamos, não é realista acreditar que usuários de catálogos de bibliotecas universitárias terão conhecimento sobre cada vocabulário utilizado para indexação em cada uma das diversas bases de dados às quais têm acesso por meio dos catálogos ou que irão realizar buscas específicas em cada uma destas bases. Especialmente, essa diversidade de vocabulários especializados, reconhecidamente necessários para atender a especificidade da representação dos conteúdos temáticos nos sistemas de recuperação da informação na internet, tende a representar uma dificuldade na realização de pesquisas interdisciplinares, interidiomáticas e interculturais.

Nesse sentido, o conceito de interoperabilidade e o objetivo de persegui-la nos sistemas de recuperação de informação visa a “[...] criar condições para que os usuários possam acessar múltiplas bases de dados, usando uma interface definida e uma linguagem de busca normalizada” conforme colocado por Moreira e Lara (2012, p. 2). Esse processo de busca, que possibilita a recuperação de resultados diversos e heterogêneos com recursos de diferentes fontes simultaneamente mesmo que as linguagens usadas para a indexação sejam diferentes, é denominado de “abordagem *one-stop*” por Zeng e Chan (2004) e diz respeito a implementação da interoperabilidade semântica entre linguagens de indexação.

A interoperabilidade semântica “[...] permite compartilhar dados diretamente, sem restrições de acesso, transformações de dados ou implementações específicas” (LLANES PADRÓN; PÁSTOR SANCHEZ, 2017, p. 391, tradução livre) porém para isso requer da representação de informação mais expressividade, maior capacidade de abstração e maior formalização. Os autores Llanes Padrón e Pástor Sanchez (2017) destacam que é importante para isso a existência de modelos conceituais, bem como, o uso de formatos de metadados padronizados para o compartilhamento harmônico de dados. A padronização e formalização necessária para este compartilhamento almejado depende, assim, do uso destes formatos comuns.

Essa constatação encontra fundamento também na norma ISO 25.964-2 (2013) que coloca que a interoperabilidade entre linguagens de indexação pode ser alcançada “[...] ao incluir mapeamentos a outros vocabulários, ao apresentar dados em formatos padrão e ao usar sistemas que têm suporte a protocolos computacionais comuns” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 7). Dessa forma, entende-se que abordar estes elementos seja essencial para compreender a aplicação e os caminhos para a interoperabilidade semântica, assim em continuidade esta seção irá contemplar os fundamentos

teóricos que permitem conceituar a interoperabilidade semântica e seus desafios, os modelos de mapeamento semântico na perspectiva teórica e da norma ISO 25.964-2 e os formatos de metadados e padrões para dados de autoridade de assunto.

Considerando a relevância da norma internacional vigente de construção de linguagens de indexação, cuja segunda parte está voltada para o estabelecimento de interoperabilidade entre linguagens, em seguida descreve-se a composição da norma ISO 25.964-2 (2013), trazendo os títulos de seus capítulos, os elementos que abordam e um exemplo significativo de seu conteúdo, quando possível.

A segunda parte da norma ISO 25964-2 (2013) compõe-se de 24 seções direcionadas à prática e aos princípios do mapeamento entre linguagens para o estabelecimento de interoperabilidade entre tesouros e outros instrumentos. Seu conteúdo abrange os objetivos da interoperabilidade, modelos estruturais para o mapeamento entre linguagens, tipos de mapeamentos, mapeamentos para recuperação da informação e, mais especificamente, a interoperabilidade entre tesouros e esquemas de classificação, taxonomias, listas de cabeçalhos de assunto, ontologias, terminologias, listas de autoridades e anéis de sinônimos (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013).

As primeiras quatro seções da norma ISO 25.964-2 (2013) apresentam elementos técnicos relativos à sua estrutura e vinculação normativa, portanto, a norma apresenta um prefácio e uma introdução e em seguida estão as seções *1 Escopo*, *2 Referências normativas*, *3 Termos e definições* e *4 Símbolos, abreviaturas e outras convenções*. Em seguida, a seção 5 representa a primeira composta de conteúdo propriamente dito.

A seção 5 *Objetivos e identificação* apresenta os objetivos da implementação da interoperabilidade entre tesouros e outras linguagens de indexação.

#### **5 Objetivos e identificação**

“Durante a recuperação da informação, o principal objetivo da interoperabilidade entre vocabulários é permitir que uma expressão formulada usando um vocabulário seja convertida para (ou suplementada por) uma expressão correspondente em um ou mais outros vocabulários. Isto se aplica se o vocabulário usa a mesma linguagem natural ou não. A expressão em questão pode ser tanto uma expressão de busca ou parte dos metadados associados a um documento. Em ambos os casos, o mapeamento é um passo chave. Se cada um dos conceitos no Vocabulário A foi mapeado ao(s) conceito(s) correspondente(s) no Vocabulário B, se torna possível trocar (ou aumentar) os termos ou identificadores que representam o conceito em cada um dos vocabulários. No nível de trabalho, a interoperabilidade é permitida pelo estabelecimento de mapeamentos entre conceitos, particularmente equivalência” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 16).

A seção 6 *Modelos estruturais para mapeamento entre vocabulários* apresenta tipos de mapeamentos, sendo eles o mapeamento de unidade estrutural, o mapeamento de link direto, o mapeamento de estrutura de hub ou eixo central e o mapeamento seletivo. No item 3.1 deste trabalho estes modelos de mapeamentos serão mais explorados.

#### **6 Modelos estruturais para mapeamento entre vocabulários**

“Os três modelos estruturais foram enumerados separadamente, mas em aplicações reais combinações deles acontecem frequentemente e as fronteiras podem ser embaçadas. No início de qualquer projeto de mapeamento é importante explicar qual modelo ou combinação de modelos será usada, com atenção particular a direção dos mapeamentos” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 20).

A seção 7 *Tipos de mapeamento* apresenta os três principais tipos de mapeamentos que são feitos entre tesouros – equivalência, hierárquicos e associativos, que se orientam aos tipos de relacionamentos existentes em tesouros. No item 3.1 deste trabalho estes tipos de mapeamentos serão mais explorados.

#### **7 Tipos de mapeamento**

“Quando relacionamentos são estabelecidos entre vocabulários, são usualmente conhecidos como mapeamentos. Em princípio, qualquer tipo de relacionamento pode ser definido e aplicado como mapeamento. Na prática, os mapeamentos que são provavelmente úteis são determinados pelos tipos de vocabulários a serem vinculados. Entre tesouros, os três tipos principais de mapeamentos a se considerar são de equivalência, hierárquicos e associativos (exatamente análogos aos relacionamentos usados internamente em qualquer tesouro). Destes o de equivalência é o tipo mais comumente necessário. Tipos adicionais de relacionamentos podem ser considerados se qualquer outro tipo de vocabulário estiver envolvido, especialmente uma ontologia” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 20).

A seção 8 *Mapeamento de equivalência* apresenta as situações possíveis de estabelecimento de equivalência entre conceitos presentes em diferentes vocabulários. Apresenta a equivalência simples, em que os dois vocabulários possuem conceitos idênticos em escopo. A equivalência composta, ou equivalência *um-para-muitos*, em que o conceito a ser mapeado foi representado por um único termo em um vocabulário e que pode ser representado por uma combinação de dois ou mais termos em outro, a equivalência composta pode ser de dois tipos, de intersecção e acumulativa. No caso da equivalência composta de intersecção, em um dos vocabulários pode haver uma situação de intersecção de dois ou mais conjuntos; na equivalência composta acumulativa um conceito complexo em um vocabulário corresponde a soma de dois ou mais conceitos em outro.

Exemplos destes tipos de mapeamentos de equivalência com uso das notações padronizadas são apresentados no quadro 5, no item 3.1.

<b>8 Mapeamento de equivalência</b>
<p>“A equivalência deve ser estabelecida quando conceitos combinados são identificados em dois ou mais vocabulários diferentes. Diferente do estabelecimento de equivalência entre dois termos em um único tesauro monolíngue, onde um deles é designado o termo preferido enquanto que o outro é designado um termo não-preferido, no contexto da equivalência inter-vocabulários o mapeamento é entre conceitos e não existe diferença de status entre os conceitos ou entre os termos preferidos ou notações que os representam” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 21).</p>

A seção 9 *Mapeamentos hierárquicos* apresenta as situações em que se podem estabelecer mapeamentos hierárquicos entre conceitos presentes em diferentes vocabulários. São apresentados os mapeamentos hierárquicos de tipo genérico, instancial e partitivo, indicando inclusive as possibilidades de distinção entre estes tipos específicos e as notações utilizadas para descrevê-los. Exemplos destes tipos de mapeamentos de equivalência com uso das notações padronizadas são apresentados no quadro 5, no item 3.1.

<b>9 Mapeamentos hierárquicos</b>
<p>“O mapeamento hierárquico pode ser estabelecido entre conceitos quando um é claramente mais amplo do que o outro. Assim como para relacionamentos e hierárquicos em um único tesauro, descritos na ISO 25.964-1:2011, 11.2, o relacionamento entre os conceitos pode ser genérico ou instancial. A situação de todo-parte pode também justificar o estabelecimento de um mapeamento hierárquico, nas circunstâncias limitadas descritas na ISO 25.964 -1:2011, 11.2.3” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 25).</p>

A seção 10 *Mapeamentos associativos* apresenta as situações possíveis de estabelecimento de relacionamentos associativos entre conceitos presentes em diferentes vocabulários.

<b>10 Mapeamentos associativos</b>
<p>“A linha divisória entre um mapeamento associativo e a equivalência inexata (ver 11.3) é má definida e subjetiva, mas este pode ser feito pragmaticamente de acordo com o contexto no qual os mapeamentos serão usados (levando em consideração, por exemplo, os interesses dos usuários, o volume de recursos relevantes a serem pesquisados, a capacidade dos sistemas de recuperação e apresentação de fazer distinções sutis)” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 25).</p>

A seção 11 *Equivalência exata, inexata e parcial* distingue entre graus que podem ser aplicados aos mapeamentos de equivalência para especificar o quanto os conceitos mapeados como equivalentes se aproximam em escopo. São diferenciados os tipos exato, em que se trata

de conceitos que podem ser alternados indistintamente; inexato, em que os conceitos podem ser tidos como equivalentes em alguns contextos, mas não em todos; e parcial, em que um dos conceitos mapeados tem um escopo parcialmente equivalente ao outro, podendo ser que um seja um pouco mais amplo do que o outro. Exemplos destes graus de mapeamentos de equivalência com uso das notações padronizadas são apresentados no quadro 5, no item 3.1.

<b>11 Equivalência exata, inexata e parcial</b>
<p>“Uma vez que a qualidade de um mapeamento específico é muitas vezes menos do que perfeita, é opcionalmente possível aplicar um marcador a um mapeamento de equivalência, indicando o grau em que o mapeamento é universalmente aplicável. Dois graus são distintos: exato (ver 11.2) e inexato (ver 11.3), que devem ser marcados usando os símbolos “=” e “~” respectivamente. A distinção deve ser aplicada apenas a casos de equivalência simples. Em casos de equivalência composta, marcadores exatos/inexatos não transmitiriam informações extras porque todos esses mapeamentos são, em certa medida, inexatos. Por outro lado, no caso de mapeamentos hierárquicos, apenas aqueles que são indiscutivelmente "exatos" devem ser aceitos como hierárquicos; os outros candidatos podem ser mais apropriadamente caracterizados como mapeamentos associativos ou equivalentes inexatos. Finalmente, no caso de mapeamentos associativos, estes abrangem tantos tipos de relacionamentos diferentes possíveis que não é significativo designá-los como exatos ou inexatos” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 26).</p>

A seção 12 *Uso de mapeamentos em recuperação da informação* expõe duas possibilidades de uso dos mapeamentos, quando os mapeamentos são utilizados nos termos de indexação de uma base de dados para recuperação de informações com termos de outro(s) vocabulário(s), ou quando os mapeamentos são utilizados nos termos de pesquisa, para que pesquisas realizadas com um vocabulário possam identificar correspondência em uma coleção indexada com outro(s) vocabulário(s).

<b>12 Uso de mapeamentos em recuperação da informação</b>
<p>“No contexto da recuperação de informação, existem dois estágios principais em que podem ser utilizados mapeamentos entre vocabulários controlados: (a) como parte do processo de indexação ou (b) no momento da pesquisa.</p> <p>NOTA Estas não são as únicas situações em que os mapeamentos podem ser usados, e as recomendações a seguir não devem ser interpretadas como excluindo outras aplicações.</p> <p>a) Quando os mapeamentos são usados no processo de indexação, os termos do índice nos metadados dos documentos indexados com o Vocabulário A são convertidos (ou aumentados por) os termos correspondentes do Vocabulário B. Isso pode ser feito rotineiramente enquanto a indexação original ocorre ou como uma conversão em massa de uma coleção completa, complementada por pequenas atualizações cada vez que os vocabulários são atualizados e/ou novos documentos são adicionados à coleção. A coleta de documentos pode então ser navegada ou pesquisada usando o Vocabulário B para o futuro indefinido, sem necessidade de mapeamento adicional.</p> <p>b) Quando os mapeamentos são aplicados aos termos de pesquisa, os metadados originais não são alterados. Para alcançar a mesma capacidade (a de usar o Vocabulário B para pesquisar uma coleção indexada com o Vocabulário A), a fonte e os vocabulários-alvo para mapeamentos precisam ser estabelecidos na direção inversa. Isso permite que consultas de pesquisa que</p>

compõem termos do Vocabulário B sejam convertidas para os termos correspondentes do Vocabulário A. Assim, os mapeamentos são incorporados ao processo de pesquisa e precisam ser aplicados toda vez que uma pesquisa é feita usando o Vocabulário B” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 28).
---

A seção 13 *Lidando com a pré-coordenação* discute as situações que podem ocorrer no caso de mapeamentos em que uma das linguagens conta com termos pré-coordenados enumerados, ou seja explicitamente pré-coordenados, e/ou sintetizados, que podem ser coordenados com base em regras providas pela linguagem. São oferecidas orientações para trabalhar com notações que dependem do contexto para serem identificadas, para situações em que o tesouro é a linguagem alvo ou a linguagem fonte do mapeamento e para situações em que o mapeamento é aplicado aos termos de indexação ou às expressões de busca, como nas situações expostas na seção 12.

<b>13 Lidando com a pré-coordenação</b>
---

“A ocorrência de conceitos, classes ou cabeçalhos pré-coordenados cria um desafio adicional para a interoperabilidade. Quando exatamente o mesmo conceito pré-coordenado ocorre em dois vocabulários diferentes, um mapeamento um-para-um pode e deve ser estabelecido. Mais comumente, a escolha de quais conceitos constituintes combinar varia de um vocabulário para outro, e isso leva a uma necessidade frequente de mapeamentos de um-para-muitos” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 32).
--

A seção 14 *Técnicas para identificação de mapeamentos candidatos* apresenta os procedimentos básicos para a identificação dos mapeamentos de relações entre conceitos presentes em diferentes linguagens de indexação. É apresentado um procedimento geral para a identificação dos mapeamentos de forma “manual”, ou seja, realizado pelo trabalho intelectual de um especialista; com assistência de um algoritmo de combinação que identifica mapeamentos candidatos que devem ser revisados por um especialista; e por análise de co-ocorrência.

<b>14 Técnicas para identificação de mapeamentos candidatos</b>
---

“Tradicionalmente, a identificação de mapeamentos é um processo intelectual. Precisa de um ou mais especialistas familiarizados com o campo de assunto relevante, fluente nos idiomas dos vocabulários a serem mapeados, e que tenham uma boa compreensão da estrutura e convenções dos vocabulários” (p. 38).
--

“Às vezes é possível automatizar o processo descrito, pelo menos em parte, usando um algoritmo correspondente. [...] Este é um campo em desenvolvimento rápido e outros métodos podem muito bem emergir. Esta parte da ISO 25964 não exclui nenhuma tecnologia emergente, mas recomenda que para mapeamentos confiáveis de uma boa qualidade, quaisquer mapeamentos candidatos que forem gerados automaticamente devem ser submetidos a revisão por um especialista” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 40).
---

A seção 15 *Gestão de dados* apresenta diretrizes para a forma de armazenamentos dos dados referentes aos mapeamentos, são indicadas as informações que devem ser armazenadas para cada tipo de modelo de mapeamento aplicado, assim como, em qual base de dados será realizado o armazenamento destes dados.

<b>15 Gestão de dados</b>
<p>“Os mapeamentos, conjuntos de mapeamentos e clusters são normalmente armazenados em um banco de dados. Para cada mapeamento devem ser especificadas as respectivas características detalhadas em 15.1. Cada mapeamento individual deve ser detalhado como um relacionamento ou como regra. O armazenamento desses mapeamentos pode ser feito usando uma tecnologia de banco de dados, como um banco de dados relacional, um banco de dados XML, um banco de dados de regras ou em RDF. Cada um desses métodos requer um esquema dedicado [...]. Ainda não há um esquema padrão que cumpra totalmente esta parte da ISO 25964, e o desenvolvimento de tal esquema não está dentro do escopo. No entanto, este é um campo em rápida evolução e os implementadores desta parte da ISO 25964 devem estar atentos aos desenvolvimentos entre as partes interessadas, por exemplo, a comunidade de usuários SKOS” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 43).</p>

A seção 16 *Apresentação de vocabulários mapeados* orienta para opções de apresentação de mapeamentos direcionadas para usuários finais, bem como, para atividades profissionais de gestão e manutenção destes mapeamentos.

<b>16 Apresentação de vocabulários mapeados</b>
<p>“Para algumas aplicações, os mapeamentos são usados principalmente por computadores e a exibição deles pode ser confusa, bem como desnecessária. Por exemplo, os humanos não precisam ver os mapeamentos se os termos que entram em um sistema de pesquisa forem convertidos automaticamente em equivalentes em outra língua. Mesmo que a mediação humana seja invocada (ver 12.2), as escolhas podem ser apresentadas ao usuário de forma que não exija que ele esteja ciente do processo de mapeamento. Por exemplo, termos mapeados do vocabulário de destino podem ser apresentados no estilo de uma nuvem de tag (comparável ao estilo mostrado em alguns aplicativos de rede social) sem designação explícita do tipo de mapeamento em cada caso” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 45).</p>

Na sequência, a norma ISO 25.964-2 (2013) inclui as seções 17 *Esquemas de classificação*, 18 *Esquemas de classificação usados para gestão de recursos*, 19 *Taxonomias*, 20 *Esquemas de cabeçalhos de assunto*, 21 *Ontologias*, 22 *Terminologias*, 23 *Listas de autoridade de nomes*, 24 *Anéis de sinônimos*. Estas seções não serão descritas como as anteriores por se aplicarem a orientações sobre o estabelecimento de mapeamentos entre tesouros e cada linguagem de indexação especificada no título da seção, dessa maneira estando além do escopo deste trabalho.



O objetivo da interoperabilidade destacado pela norma ISO 25.964-2 (2013) é o de possibilitar o uso de expressões de busca formuladas com uma linguagem automaticamente convertida a uma ou mais outras linguagens para recuperação de informação. Para atender a esse objetivo os mapeamentos são o elemento chave, especialmente o mapeamento de relações semânticas.

Durante a recuperação da informação, o principal objetivo da interoperabilidade entre vocabulários é permitir que uma expressão formulada usando um vocabulário seja convertida para (ou suplementada por) uma expressão correspondente em um ou mais outros vocabulários. No nível de trabalho, a interoperabilidade é permitida pelo estabelecimento de mapeamentos inter-conceitos, particularmente equivalência [...]. (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 16).

Outra possibilidade da interoperabilidade semântica é o próprio aprimoramento das linguagens, como a possibilidade de criação de novas linguagens a partir da fusão de linguagens existentes e o uso de uma parte de uma linguagem para desdobrar outra linguagem menos desenvolvida naquele segmento.

O estabelecimento da interoperabilidade semântica, entretanto, envolve uma quantidade de aspectos que podem tornar-se problemáticos. Lucarelli e Viti (2015) elencaram por exemplo: a estrutura conceitual das ferramentas e linguagens que serão conectadas (principalmente se tratarem-se de linguagens com coordenação diferente); a compatibilidade entre domínios específicos e gerais; o nível de especificidade das linguagens; o significado dos termos, quanto a polissemia e ambiguidades; as convenções gramaticais e a sintaxe da linguagem ao se lidar com usos regionais/nacionais; o percentual de ligação entre linguagem fonte e linguagem alvo e o tipo de equivalências encontradas.

Diante do exposto, considera-se a necessidade de compreender os mapeamentos semânticos e os métodos de realização destes mapeamentos entre linguagens de indexação, que será discutido na sequência.

### **3.1 Modelos de mapeamento semântico**

O principal mecanismo apresentado na literatura como meio para alcançar interoperabilidade semântica é o denominado mapeamento que significa em um nível elementar “o processo de mapear vocabulários basicamente consiste em estabelecer equivalência entre os termos em diferentes vocabulários controlados ou entre termos verbais e números de classificação”, como explicam Zeng e Chan (p. 383, 2004, tradução livre).

A norma ISO 25.964-2 (2013) indica que o objetivo de realizar mapeamentos é indicar os relacionamentos entre conceitos presentes em diferentes vocabulários. O processo realizado no mapeamento transforma o conjunto de dados da linguagem de indexação para que esta se torne interoperável e dessa forma permite que seja processada por diferentes aplicações na web (LEIVA MEDEROS et al., 2017).

Nesse sentido o mapeamento é útil para melhorar os resultados de pesquisas, habilitando o sistema a pesquisar em diferentes bases de dados. Lucarelli e Viti (2015) explicam que é o estabelecimento de *crosswalks* entre conceitos presentes em duas ou mais linguagens que dão suporte à esta pesquisa heterogênea, garantindo a integridade dos resultados mesmo que a indexação seja realizada com diferentes linguagens em cada base de dados. Os *crosswalks* segundo a norma ISO 25.964-2 (2013, p. 4), são tabelas de mapeamentos entre conceitos que envolvem dois ou mais vocabulários estruturados que permitem aos motores de busca realizarem pesquisas entre bases de dados heterogêneas.

Em estudo sobre as relações semânticas, Mazzochi (2017) aponta os mapeamentos como uma opção flexível para o estabelecimento da interoperabilidade semântica.

[...] alcançar a interoperabilidade semântica, requer uma compreensão comum do significado do conteúdo que é trocado entre sistemas diferentes. Estabelecer relacionamentos (ou mapeamento) entre os termos/conceitos de tesouros diferentes está envolvido aqui [...], e o papel desempenhado pelas relações fundamentais de tesouros, que proporcionam uma base partilhada, é crucial (MAZZOCHI, 2017, p. 378).

**Quadro 4** - Elementos legíveis por humanos usados para representar conceitos em mapeamentos

Tipo de Vocabulário	Conceitos representado por:
Tesouros	Termos preferidos
Esquemas de classificação	Notações
Taxonomias	Notações ou rótulos de categorias
Esquemas de cabeçalho de assunto	Cabeçalhos
Listas de autoridades de nome	Nomes
Ontologias	Rótulos
Terminologias	Termos ou outro tipo de designação

**Fonte:** Traduzido de ISO 25964-2 (2013)

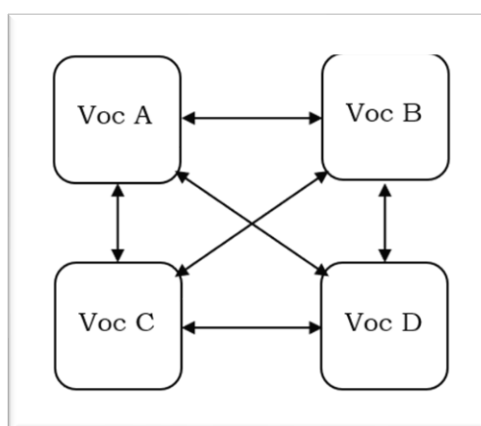
O Quadro 4 apresenta um guia para a identificação dos elementos das linguagens de indexação, os conceitos, a serem mapeados segundo a norma ISO 25.964-2 (2013), ressaltando a diversidade terminológica para indicá-los em cada tipologia de linguagem. Estas nomenclaturas são utilizadas quando os mapeamentos são direcionados para leitores humanos. Porém, quando os mapeamentos são direcionados a manipulação por computadores, a indicação

da norma é de que os conceitos devem ser representados por identificadores únicos e persistentes, tais como o *Uniform Resource Identifier* (Identificador Uniforme de Recursos - URI).

A norma ISO 25964-2 (2013) identifica três modelos estruturais de mapeamento:

- De unidade estrutural – os conceitos e as estruturas relacionais e ambas as linguagens de indexação são equivalentes;
- De link direto – no qual todas as linguagens de indexação se interligam entre si;
- De hub, ou eixo central – no qual mapeiam-se as linguagens de indexação em ambas as direções, a partir de um que funciona como núcleo do sistema de vocabulários estruturados.

**Figura 3** - Modelo 2 (link direto) como aplicado a quatro vocabulários

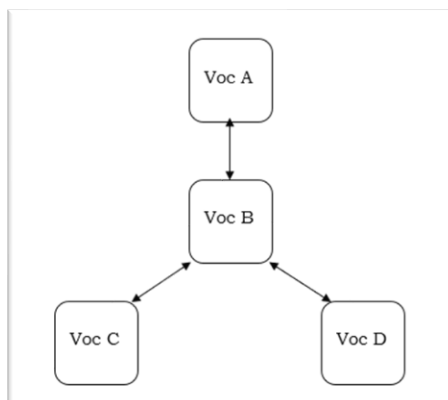


**Fonte:** ISO 25.964-2 (2013, p. 17)

A Figura 3 ilustra os mapeamentos que são necessários para lidar com quatro vocabulários, usando o modelo de link direto. As setas duplas na figura indicam a realização de mapeamento bidirecional, em que foram realizados os mapeamentos considerando cada vocabulário como fonte e posteriormente como alvo dos mapeamentos.

A norma também prevê a possibilidade de um mapeamento seletivo, entre partes dos vocabulários, sem buscar um mapeamento completo, mas com objetivo mais prático de aproximação, voltado apenas para os conceitos que têm sido usados ou têm probabilidade de que sejam usados na linguagem de indexação alvo.

A Figura 4 mostra o modelo 3, no qual o vocabulário B serve como um *hub* ou núcleo e os vocabulários A, C e D como “satélites”.

**Figura 4** - Modelo 3 (modelo de hub) como aplicado a quatro vocabulários

**Fonte:** ISO 25.964-2 (2013, p. 18)

Quanto às relações semânticas, são previstos o mapeamento de relações de equivalência, hierárquicas e associativas, fundamentalmente. O Quadro 5 apresenta estas relações com exemplos e suas respectivas siglas e, além da tradução do original de Garcia Marco (2017), alguns exemplos foram adaptados no intuito de estabelecer uma melhor compreensão em português.

**Quadro 5** - Tipos de mapeamentos encontrados na ISO 25.964-2 (2013)

Tipos	Subtipos	Etiqueta	Exemplos
Equivalência simples		EQ	Smartphones EQ Telefones inteligentes
	Exata	=EQ	Manjerição =EQ Ocimum basilicum
	Inexata	~EQ	Ensino ~EQ Educação
Equivalência composta	De intersecção	+	Touros EQ Gado bovino + Machos
	Acumulativa	EQ	Combustíveis fósseis   carvão   petróleo   gás natural
Hierárquico	Hiperonímica	MG	Canais MG Vias de comunicação aquáticas
	Hoponímica	ME	Vias de comunicação aquáticas ME Canais
Associativo		MR	Plataformas docentes MR E-learning

**Fonte:** Adaptado de GARCIA MARCO, 2017.

As siglas apresentadas no Quadro 5 estão estabelecidas na norma ISO 25964-2 (2013) e são distintivas dos tipos de equivalência entre conceitos que podem ser inventariadas nos mapeamentos entre SOC, sendo da seguinte forma para os relacionamentos de equivalência: Equivalência simples (EQ), que pode ser do subtipo exata (=EQ) ou inexata (~EQ); Equivalência Composta De intersecção (EQ+); Equivalência Composta Acumulativa (EQ |).

Quanto aos relacionamentos hierárquicos as siglas MG e ME (adaptações para o português de BM e NM) são acrônimos para as denominações *Broader Mapping* (Mapeamento

Geral) e *Narrower Mapping* (Mapeamento Específico), respectivamente. Da mesma forma, para os relacionamentos associativos a sigla MR (adaptação para o português de RM) refere-se à *Related Mapping* (Mapeamento Relacionado). Estas siglas foram estabelecidas para assemelhares-se àquelas relativas aos relacionamentos hierárquicos e associativos existentes tradicionalmente em tesouros, de modo que possam ser facilmente identificadas em respostas de buscas com linguagens de indexação mapeadas. O quadro 6 apresenta as notações estabelecidas pela ISO 25.964-2 para relacionamentos mapeados, adaptadas para o português e no idioma original, acompanhadas do tipo de mapeamento que indicam.

**Quadro 6** - Notações estabelecidas da ISO 25.964-2 (2013) para mapeamentos entre linguagens

Notação	Descrição	Notação Original ISO (inglês)
MGG	Mapeamento Geral Genérico	BMG
MEG	Mapeamento Específico Genérico	NMG
MGP	Mapeamento Geral Partitivo	BMP
MEP	Mapeamento Específico Partitivo	NMP
MGI	Mapeamento Geral Instancial	BMI
MEI	Mapeamento Específico Instancial	NMI
MR	Mapeamento Relacionado	RM
EQ	Equivalência simples	EQ
=EQ	Equivalência simples exata	=EQ
~EQ	Equivalência simples inexata	~EQ
EQ +	Equivalência composta de intersecção	EQ +
EQ	Equivalência composta acumulativa	EQ

**Fonte:** Elaboração própria.

Diante do apresentado, concordamos com Mazzochi (2017) que afirma que a interoperabilidade semântica alcançada por meio dos mapeamentos entre linguagens de indexação é uma opção flexível para atender a necessidades particulares e que permitem a coexistência de sistemas diferentes, respeitando as características e objetivos para os quais foram construídos ao passo que permite uma integração em um nível macro.

### 3.2 Formatos e padrões de metadados para dados de autoridade

O uso de uma linguagem de indexação adequada e atualizada para a representação dos conceitos significativos de um documento no processo de indexação é essencial para a garantia da recuperação da informação em um sistema. A construção de uma linguagem deste porte é um processo complexo que requer que se respeitem os princípios teóricos estabelecidos na literatura de Organização do Conhecimento (BARITE, 2011; CAMPOS, 2001; CURRÁS, 2005; MOREIRA; MOURA, 2006) e que seja norteado pelas diretivas estabelecidas nas normas

internacionais de construção (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011).

Entretanto, com os desenvolvimentos tecnológicos ocorridos nas últimas décadas, este processo de construção de linguagens de indexação tem sido facilitado pelo desenvolvimento de softwares que já podem ser interligados às bases de dados bibliográficos e aos catálogos de acesso público. Esse recurso representa não apenas uma vantagem ao profissional de informação, que constrói a linguagem e a utiliza nos processos de tratamento temático, mas também aos usuários que podem ter à sua disposição uma ferramenta integrada que lhe permita efetivamente realizar suas pesquisas com a mediação da linguagem de indexação, realizando buscas de modo mais intuitivo e, assim, obtendo resultados mais condizentes com seus interesses.

Porém, para que estes desenvolvimentos ocorram, um elemento se coloca como chave no processo de construção de linguagens de indexação, o registro de autoridade de assunto. O registro de autoridade de assunto é o responsável pelo controle de vocabulário automático no sistema, ao incluir para cada termo a descrição do seu significado e das relações com outros termos, sejam hierárquicos ou associativos e, com isso realizar adequações e correções automaticamente referentes a cada termo quando de seu uso no campo de assunto do registro bibliográfico no formato MARC21.

Para compreender a função do registro de autoridade de assunto na construção de linguagens de indexação, é preciso recuperar o seu contexto de desenvolvimento no âmbito da catalogação descritiva e de assunto.

O processo de criação de registros de autoridades, pautados nos princípios do controle de autoridade, remonta aos primórdios da padronização do trabalho bibliotecário, em que eram registrados em fichas construídas manualmente. De acordo com Campos (2003) o controle de autoridade tem como objetivos a validação do trabalho do catalogador, percebido pela qualidade do catálogo, por meio da criação e manutenção de registros de autoridade e a garantia do acesso à informação consistente e confiável.

Esse processo de controle de autoridade encontrou solo fértil para seu desenvolvimento com o avanço do uso das tecnologias de informação nos ambientes de bibliotecas. Na visão de Campos (2003), o registro de autoridade tornou-se elemento obrigatório nos catálogos online, pois se conecta aos registros bibliográficos e é a partir das autoridades que se dá o processo de consulta à base de dados do catálogo, assim como, da disposição dos resultados pertinentes.

A preocupação com a normalização e a orientação à criação de registros de autoridade padronizados tem estado presente nas discussões da *International Federation of Library*

*Associations and Institutions* (IFLA) com a constituição de grupos de trabalho voltados a estes temas, o que teve como resultados a publicação das *Guidelines for Authority and References Entries* (GARE) em 1984, atualizada em 2001 como *Guidelines for Authority Records and References* (GARR), e refletida na composição dos formatos de metadados de autoridade, como o *United International Machine Readable Cataloging* (UNIMARC) (LLANES PADRÓN; BASTOS; FUJITA, 2014).

Nesse contexto, hoje os pontos de acesso por assuntos em geral são construídos a partir da criação de um registro de autoridade de assunto por meio de um arquivo de autoridade geralmente em formato *Machine Readable Cataloging* - MARC21.

Dentre os avanços ocorridos no campo da construção de registro de autoridade de assunto deve-se dar destaque à publicação em 2010 do modelo conceitual *Functional Requirements for Subject Authority Data* (Requisitos Funcionais para Dados de Autoridade Assunto – FRSAD), que vem a fazer parte do que se tem denominado “Família FRBR”, já composta pelo próprio *Functional Requirements for Bibliographic Records* (Requisitos Funcionais para Registros Bibliográficos - FRBR) e cuja primeira expansão, no que tange às formas autorizadas de nome, havia tomado forma no modelo conceitual *Functional Requirements for Authority Data* (Requisitos Funcionais para Dados de Autoridade – FRAD).

Nesse cenário, o modelo FRBR define três grupos de entidades, sendo que o primeiro grupo trata daquilo que é descrito no registro bibliográfico, ou seja, os próprios documentos, e é abordado no FRBR; o segundo grupo aborda as entidades responsáveis pelo conteúdo do documento e é abordado no modelo FRAD; o terceiro grupo, composto pelas entidades *thema* e *nomen*, é abordado no modelo FRSAD, enfocando as tarefas de representação temática da informação (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2010).

O modelo conceitual FRSAD pode ser considerado uma tendência na representação dos assuntos de documentos em catálogos, tendo como objetivos “prover uma compreensão clara e compartilhada das informações que os dados de autoridade assunto devem fornecer para corresponder às necessidades dos usuários” e o uso e compartilhamento de dados de autoridade de assunto (MELO; BRASCHER, 2014, p. 109).

Segundo MARC21 (2005), o ato de registrar uma autoridade consiste no estabelecimento de formas padronizadas de nomes (pessoais, corporativos, eventos, jurisdição, título uniforme, combinações nome/título) e assuntos (termos tópicos, nomes geográficos, nomes com subdivisão de assunto, termos e subdivisão de assunto), sendo ambos usados como pontos de acesso, ao padronizarem as entradas principais e secundárias no registro bibliográfico

com o termo nome para entrada principal e secundária e o termo assunto no catálogo online construído por meio de software de gestão e manutenção de catálogos.

Os pontos de acesso por assunto “[...] são constituídos por termos (palavra ou grupo de palavras) que representam o documento. Por eles serem padronizados permitem que a representação da informação seja feita de maneira consistente” (MARTINHO, 2010, p. 57). Ou seja, por meio de um registro de autoridade de assunto pode-se estabelecer a forma padronizada dos pontos de acesso por assunto selecionados na indexação para representação no registro bibliográfico.

O registro de autoridade de assunto é composto por indicações de relações e notas de escopo, e tem quatro funções principais, fornecer o termo de assunto autorizado ou referir ao termo autorizado, indicar se um termo pode ou não ser subdividido geograficamente, fornecer uma nota de escopo, quando houver necessidade de explicar a cobertura a que o termo se aplica, e indicar relações entre termos (FUJITA; SANTOS, 2016).

A função mais importante de um registro de autoridade em um sistema de informação é estabelecer a forma autorizada do assunto principal de um documento, e também as formas não autorizadas, ou seja, que não são estabelecidas como assunto principal, mas que podem ser considerados como termos equivalentes e fazer a conexão entre a busca do usuário e a recuperação da informação.

A padronização e formalização necessária para o compartilhamento almejado depende do uso destes formatos comuns. O desenvolvimento do padrão SKOS (*Simple Knowledge Organization Systems*) pelo consórcio W3C em conjunto com a equipe de desenvolvimento da norma ISO 25964 destaca-se neste contexto.

O padrão SKOS é um formato para representar linguagens de indexação em *Resource Description Framework* (RDF) e tem sido amplamente utilizado em experiências envolvendo esforços para agregar interoperabilidade semântica aos vocabulários, como é o caso de Leiva Mederos et al. (2017). Martínez González e Alvite Díez (2014) em estudo de avaliação de ferramentas para construção e gerenciamento de linguagens de indexação enfatizam que RDF e SKOS são padrões da Web Semântica que “[...] permitirão aos tesouros alcançar um lugar de sucesso na interoperabilidade da informação, o que é muito importante em sistemas de informação na web” (MARTÍNEZ GONZÁLEZ; ALVITE DÍEZ, 2014, p. 712, tradução livre).

O uso deste padrão permite distinguir entre diferentes conceitos e os rótulos pelos quais serão designados (termos, notações, *tags*), isto porque o conceito é uma entidade formal representada por um URI que independe de sua forma em linguagem natural (LUCARELLI; VITI, 2015, p. 417).



Diante do exposto acerca da interoperabilidade semântica e dos mapeamentos semânticos, pode-se entender que um projeto que objetive o estabelecimento de interoperabilidade semântica entre linguagens de indexação necessite preencher os seguintes elementos: a definição de uma equipe de trabalho com expertise nos processos relativos à construção e gestão de linguagens de indexação, com atribuição de responsabilidades aos membros desta equipe; a definição dos objetivos da linguagem de indexação e o público-alvo desta linguagem; uma avaliação dos recursos humanos, materiais e de vocabulário disponíveis; a definição do software de gestão do tesouro que será utilizado; a definição do modelo de mapeamento que será adotado para o estabelecimento de relações e a direção dos mapeamentos; apresentar as declarações de mapeamento indicando o tipo de mapeamento entre os conceitos com o uso das notações padronizadas e o uso de identificadores únicos e persistentes, tais como o URI, especialmente para que o projeto possa ser realizado com assistência de um algoritmo de combinação e para o uso dos mapeamentos para recuperação da informação.

Exemplos de experiências com projetos voltados à construção e manutenção de linguagens de indexação visando a interoperabilidade semântica em bibliotecas ainda não são frequentes, mas alguns podem ser identificados na literatura corrente. Como exemplo de experiência internacional, Lucarelli e Viti (2015) desenvolveram na Biblioteca Nacional Central de Florença - Itália um projeto de interoperabilidade semântica entre o tesouro *Nuovo Soggettario*, amplamente utilizado por bibliotecas e instituições culturais italianas, e a LSCH. As autoras explicam que desde 2010 os metadados do *Nuovo Soggettario* estão publicados em formato SKOS/RDF o que as permitiu desenvolver o projeto, que implementou um mapeamento parcial estabelecendo relações de equivalência entre termos preferidos mapeados nas duas linguagens de indexação.

Como exemplo de experiência nacional, Fujita et al. (2018) desenvolveram projeto de elaboração de uma linguagem de indexação para as bibliotecas universitárias da Rede Unesp com base em compatibilização de três linguagens existentes, a Terminologia da Biblioteca Nacional, sua fonte a LCSH e o Vocabulário Controlado da USP, com a realização de mapeamento manual.

Considerando os elementos apresentados nestes dois capítulos que formalizam as bases teóricas da pesquisa, no próximo capítulo deste trabalho aborda-se os procedimentos metodológicos que foram aplicados, em que buscamos identificar projetos de interoperabilidade semântica em bibliotecas universitárias.

## 4 METODOLOGIA

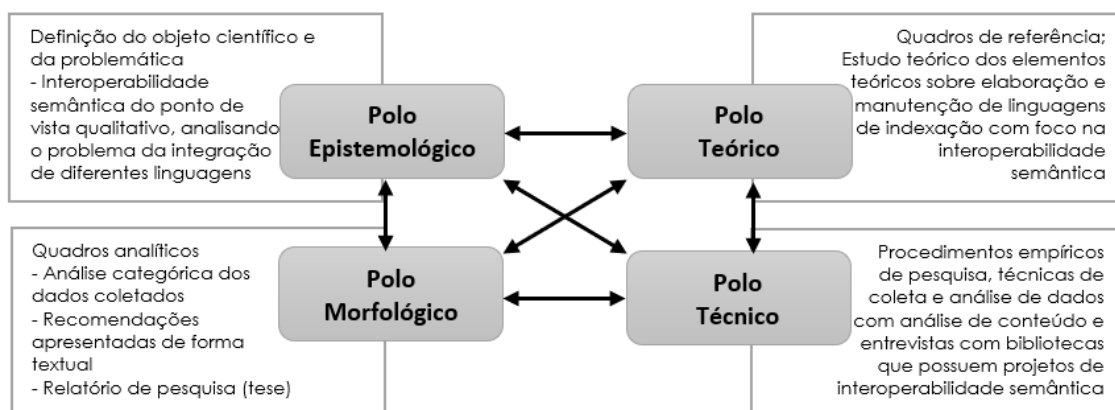
Para a realização desta investigação definiu-se como abordagem o uso do método quadripolar de pesquisa, descrito por Bruyne, Herman e Schoutheete (1991), que indicam a estruturação da pesquisa nas Ciências Sociais em quatro polos: polo epistemológico, polo teórico, polo técnico e polo morfológico.

O método quadripolar foi inicialmente proposto pelos pesquisadores belgas Paul De Bruyne, Jacques Herman e Marc de Schoutheete em 1974 e seu uso na Ciência da Informação foi primeiramente indicado por Silva e Ribeiro (2002) que o consideraram apto a atender às pesquisas com foco nos fenômenos relativos à informação, evidenciando a abordagem qualitativa que é própria das Ciências Sociais Aplicadas, tais como a Ciência da Informação.

O método propõe que a pesquisa se desenvolva articulando-se os quatro polos entre si de forma dinâmica ao longo do desenvolvimento da investigação, o que coloca o método quadripolar no nível de uma metodologia geral que orienta o esquadramento de toda a pesquisa e não apenas um guia técnico ou instrumental.

A Figura 5 apresenta uma visualização da sistematização desta interação entre os polos de pesquisa propostos no método quadripolar, que serão detalhados em seguida.

**Figura 5** - Sistematização do método quadripolar e a interação entre os polos



**Fonte:** Elaboração própria com base em Bruyne, Herman e Schoutheete (1991)

O polo epistemológico, é o polo de reflexão sobre a pesquisa e sobre o processo de produção de conhecimento. Segundo os autores, Bruyne, Herman e Schoutheete (1991, p. 42), a função da epistemologia é a de vigilância sobre as condições de objetividade em relação à delimitação do objeto científico e da problemática de pesquisa, bem como, em relação aos procedimentos de observação e experimentação dos fatos e a compreensão das teorias e suas

relações com estes fatos, com o objetivo de estabelecer regras para o procedimento científico. Esta compreensão de epistemologia a coloca enquanto um polo essencial da pesquisa determinando uma visão da ciência não como um produto, mas como um processo vivo.

No polo epistemológico, que busca compreender o processo de construção de conhecimento, consideramos relevante a abordagem da interoperabilidade semântica enquanto objeto de pesquisa do ponto de vista qualitativo, com foco no elemento humano, ou seja, ponderando os aspectos normativos, metodológicos, políticos, contextuais e sociais, para além de apenas tratar dos elementos técnicos e tecnológicos envolvidos com o tema.

Para isso, o problema de pesquisa formulado foi, a disseminação de vocabulários com níveis de controle diversos construídos individualmente por bibliotecas universitárias brasileiras como fator de isolamento, por não atenderem às normativas internacionais e assim não permitirem o alcance da interoperabilidade semântica com outros vocabulários.

O polo teórico é o local da formulação de sistemas conceituais e, embora se materialize pela construção de um quadro conceitual, segundo os autores, Bruyne, Herman e Schoutheete (1991), a teoria manifesta-se em todo o processo da pesquisa empírica. Desse modo a teoria é parte do processo metodológico e exerce função de ruptura com as noções do senso comum.

Dessa forma, a teorização perpassa todo o processo de pesquisa iniciando-se quando a pesquisa começa e funcionando como um polo interno da pesquisa, tendo como propósito intrínseco realizar uma ligação entre um momento denominado de plano de descoberta, onde as teorias e hipóteses são formuladas, ao momento denominado de plano de prova, no qual estas teorias e hipóteses são aceitas ou rejeitadas (BRUYNE; HERMAN; SCHOUTHEETE, 1991).

O polo teórico é o local do embasamento que permitirá a realização da pesquisa e se viabilizará pela construção de uma fundamentação teórica relativa à linguagem de indexação e interoperabilidade semântica e dentro do quadro conceitual da Organização e Representação do Conhecimento em Ciência da Informação.

A hipótese colocada nesta pesquisa entende que para que bibliotecas universitárias possam construir e manter atualizadas linguagens de indexação que atendam às normativas internacionais necessárias à integração dos diferentes sistemas de representação de informação a interoperabilidade semântica tem importância estratégica.

O quadro 7 sistematiza a descrição dos procedimentos metodológicos da pesquisa, relacionando-os aos polos estabelecidos no método quadripolar e aos objetivos de pesquisa.

**Quadro 7** - Descrição dos procedimentos metodológicos e integração com polos da pesquisa

<b>Polos</b>	<b>Fases</b>	<b>Descrição dos procedimentos metodológicos</b>	<b>Objetivos</b>
<b>Epistemológico</b>	Fase 1	a. Construção do objeto de estudo e delineamento das premissas da pesquisa e da problemática referente ao tema b. Construção da hipótese e da tese da pesquisa	<b>Objetivo Geral</b> Oferecer subsídios teóricos e metodológicos à pesquisa sobre a interoperabilidade semântica de linguagens de indexação no contexto teórico da Organização e Representação do Conhecimento.
<b>Teórico</b>	Fase 2	c. Revisão de literatura	<b>Objetivos específicos</b> 1. Realizar estudo teórico e metodológico com revisão da literatura nacional e internacional sobre elaboração e manutenção de linguagens de indexação com foco na interoperabilidade semântica
<b>Técnico</b>	Fase 3 Etapa 3A	d. Análise de conteúdo da norma internacional de construção de tesouros e. Selecionar amostra de bibliotecas com projetos de interoperabilidade semântica, preparar roteiro de entrevista semiestruturada f. Redação prévia de recomendações para construção e manutenção de linguagem de indexação interoperável	2. Analisar o processo de elaboração de linguagens de indexação sob o aspecto das normas internacionais focalizando a interoperabilidade semântica;
	Fase 3 Etapa 3B	g. Entrevistas semiestruturadas com participantes de projetos de linguagem de indexação validação das recomendações para construção e manutenção de linguagem de indexação interoperável por meio de.	3. Identificar e analisar projetos de interoperabilidade semântica de linguagens de indexação
<b>Morfológico</b>	Fase 4	h. Redação final das recomendações para a construção de linguagens de indexação em bibliotecas i. Redação do relatório final de pesquisa	4. Elaborar recomendações para a construção de linguagens com interoperabilidade semântica

Fonte: Elaboração própria.

#### 4.1 O polo técnico

O polo técnico, é delimitado por Bruyne, Herman e Schoutheete (1991) como aquele que

[...] trata dos procedimentos de coleta das informações, das transformações destas últimas em dados pertinentes à problemática geral; tem a função de circunscrever os “fatos” em sistemas significantes, por protocolos de evidênciação experimental desses dados empíricos (BRUYNE, HERMAN; SCHOUTHEETE, 1991, p. 201).

Segundo os autores, as técnicas de coleta propiciam transformações nas informações que são coletadas e neste sentido existe uma interação com o polo epistemológico, pois realizam-se tais transformações de modo a relacionar estas informações com a hipótese dada e a abordagem ao problema (BRUYNE; HERMAN; SCHOUTHEETE, 1991, p. 204-206).

Quanto às técnicas de coleta de dados propriamente ditas, os autores abordam especificamente três modos de coleta, a pesquisa por entrevistas e questionários, as pesquisas de observações direta e participante e as análises documentais (BRUYNE; HERMAN; SCHOUTHEETE, p. 210-214).

O polo técnico aborda a realização da pesquisa empírica que se colocará “em confronto com a teoria que foi suscitada” (SILVA, 2014, p. 36). São utilizadas como técnicas de coleta de dados, além da pesquisa bibliográfica essencial para o polo teórico, a análise de conteúdo e a entrevista semiestruturada que, entre outras, foram indicadas como pertinentes à aplicação do método quadripolar na Ciência da Informação por Silva (2014).

O método de Análise de Conteúdo é na verdade um conjunto de instrumentos metodológicos aplicáveis à descrição do conteúdo comunicacional, seja ele produzido na comunicação verbal ou por meio de textos escritos, de forma sistemática e orientada à realização de inferências dedutivas a respeito dos dados analisados (BARDIN, 2011). Trata-se de um método qualitativo e próprio das Ciências Humanas e Sociais.

Amorim e Café (2006) definem a Análise de Conteúdo como

[...] um método que abrange de forma sistemática o tratamento e análise dos dados, compreendendo etapas que vão desde o levantamento e seleção de bibliografias e estruturação do corpus de análise, até a etapa de tratamento de resultados, construção de inferências e interpretação (AMORIM, CAFÉ, 2016, p. 10).

A realização da Análise de Conteúdo segundo Bardin (2011, p. 125) prevê três fases de execução: a pré-análise; a exploração do material; e o tratamento dos resultados, a inferência e a interpretação.

Na fase de pré-análise, como o próprio nome sugere, são realizadas as aproximações iniciais ao desenvolvimento das operações de pesquisa de modo que se realize a organização e sistematização de um plano de análise. Bardin (2011) orienta que sejam realizadas nesta etapa a definição dos objetivos da análise, a escolha dos documentos que irão compor o *corpus* de análise, a preparação deste material para a análise, a elaboração de índices e indicadores e a leitura flutuante.

Tais atividades estão totalmente interconectadas e não necessitam ser realizadas em uma ordem específica. Mas destaca-se que entre estas atividades a chamada leitura flutuante possui especial importância pois intermedia a realização da preparação do material, bem como, de qualquer seleção deste material para melhor adequar o *corpus* de análise e de testagem das técnicas de análise.

A fase de exploração do material é a fase da análise propriamente dita, que se dá de acordo com as regras previamente formuladas. Segundo Amorim e Café (2016) nesta fase ocorre a “[...] extração e a descrição dos elementos sobre os quais se podem constituir categorias” (AMORIM, CAFÉ, 2016, p. 10). Dessa forma, pode ser realizado um quadro de categorias, em que se aborde a frequência com que temas e subtemas aparecem nos conteúdos analisados.

Na etapa de tratamento dos resultados e interpretações realiza-se a síntese e seleção dos resultados, as inferências de acordo com as hipóteses de pesquisa e a interpretação destes resultados.

A técnica de análise de conteúdo foi utilizada para análise das normas internacionais de construção de linguagens de indexação. A entrevista semiestruturada, foi utilizada para aprofundamento das descobertas alcançadas com a análise da documentação. A opção por esta técnica se deu no intuito de obter uma aproximação maior com o objeto de pesquisa, buscando interagir com os responsáveis pelas linguagens analisadas.

A técnica de entrevistas, de acordo com Bruyne, Herman e Schoutheete (1991) pode ser realizada nas modalidades estruturada, em que há um protocolo fixo; livre; centralizada num tema particular; informal e contínua; painel, em que são realizadas repetidas entrevistas e; em profundidade indireta.

As informações que são coletadas podem ser dos tipos fatos observados ou opiniões; mudanças de atitudes; evolução dos fenômenos; significação das respostas e; conteúdo latente. A seleção dos informantes pode ser feita por amostragem, por seleção de respondentes considerados representativos ou por seleção de pessoas consideradas competentes (*keyinformants*).

Os autores explicam que esta técnica pode apresentar fraquezas que podem levar à resultados indesejados, como inconsistências de comunicação – tais como subjetividade da resposta; inadequação dos conceitos com o real; incompreensões; estado de informação aleatório – ou inconsistências comportamentais – tais como mecanismos de defesa do respondente, como a fuga ou recusa, e diferenças entre as declarações e o comportamento apresentado. Porém, a técnica de entrevistas conta com benefícios, como a quantidade e a

qualidade ampliada das informações, podendo serem expostos problemas mais complexos e a flexibilidade na aplicação.

#### *4.1.1 Elaboração do protocolo de entrevista e realização de entrevista semiestruturada*

Diante do exposto, optou-se por realizar a entrevista semiestruturada, com protocolo definido, porém com possibilidade de inclusão de questões específicas relativas às particularidades conhecidas de cada instituição e com seleção por respondentes considerados representativos.

Com a aplicação da análise de conteúdo de Bardin (2011) às duas partes da norma ISO 25.964 (2011, 2013), que foram apresentadas nos capítulos 2 e 3, foi possível identificar pontos decisivos a respeito dos procedimentos de construção de linguagens de indexação e estabelecimento de interoperabilidade semântica entre linguagens, estes pontos foram determinantes para a elaboração do protocolo de entrevista que, tendo em vista a sua análise, foi organizado em 4 seções compostas por perguntas assemelhadas em sua temática, sendo estas:

- Identificação do respondente e instituição
- Linguagem de indexação
- Formatos de metadados padrão/ Protocolos computacionais
- Mapeamentos semânticos

No Quadro 8 apresentamos uma sistematização que inclui as seções do protocolo de entrevista semiestruturada contendo as perguntas padrão que as compõem, que foram direcionadas a todos os respondentes, juntamente com o objetivo de análise destas perguntas. Ressalte-se que, tratando-se de um protocolo direcionado à realização de entrevista semiestruturada, à estas questões somam-se questões específicas direcionadas a cada respondente de acordo com o andamento da entrevista.

**Quadro 8** - Sistematização da relação entre perguntas e categorias do protocolo de entrevistas

<b>Seções</b>	<b>Descrição das perguntas do protocolo de entrevista e objetivo das seções</b>
<b>Identificação do respondente e instituição</b>	Foram elaboradas duas questões a fim de identificação do respondente quanto às atividades desempenhadas na instituição e ambientação do respondente com a entrevista. <ol style="list-style-type: none"> <li>1. Identificação da biblioteca e respondente:</li> <li>2. Respondente desempenha qual função na biblioteca?</li> </ol>
<b>Linguagem de indexação</b>	Foram elaboradas duas questões voltadas a verificação do estado do desenvolvimento do projeto de interoperabilidade existente na biblioteca, conforme atendimento dos critérios de seleção. <ol style="list-style-type: none"> <li>3. A biblioteca desenvolve ou desenvolveu projeto construção ou manutenção/atualização de linguagem de indexação com interoperabilidade semântica? Discorra.</li> <li>4. Se sim. Qual software está sendo utilizado para desenvolver a linguagem?</li> </ol>
<b>Formatos de metadados padrão/ Protocolos computacionais</b>	Foi elaborada uma questão relativa aos formatos e padrões de metadados para o estabelecimento de interoperabilidade semântica entre linguagens de indexação. <ol style="list-style-type: none"> <li>5. Qual formato de metadados foi adotado para desenvolver a linguagem?</li> </ol>
<b>Mapeamentos semânticos</b>	Foram elaboradas cinco questões com objetivo de identificar o estágio de desenvolvimento do projeto de interoperabilidade existente na biblioteca e a compreensão dos respondentes acerca das dificuldades encontradas neste processo. <ol style="list-style-type: none"> <li>6. Está trabalhando com linguagens que possuem modos de coordenação diferentes? Ex. PréC- CDD/CDU e PósC – LCSH, tesouros... S/N</li> <li>7. Houve dificuldades? Como a questão da coordenação está sendo resolvida?</li> <li>8. Está trabalhando com linguagens que possuem níveis de especificidade diferentes? Ex. Geral: TBN, LCSH e Específicas: MeSH, DeCS, Tesouros de áreas, Agrovoc etc. S/N</li> <li>9. Houve dificuldades? Como a questão de especificidade das linguagens fonte está sendo resolvida?</li> <li>10. Qual outra dificuldade foi encontrada durante o processo de desenvolvimento/atualização da linguagem com relação à interoperabilidade semântica?</li> </ol>

Fonte: elaboração própria

De acordo com as orientações de Bardin (2011) as entrevistas foram registradas por meio de gravação de áudio e, posteriormente, integralmente transcritas (Apêndices A, B e C) para a realização da análise das respostas. A cada entrevista foi atribuído um código numérico sequencial para sua identificação. As entrevistas foram realizadas à distância, via Skype, meio pelo qual também foi realizada a gravação. Apenas uma entrevista não pode ser gravada pois, por questões internas, os entrevistados não possuíam acesso ao Skype, a entrevista foi realizada



por um sistema de videochamadas hospedado em um site externo, o que não permitiu o registro da entrevista. A transcrição disponível no Anexo C foi realizada com base no registro por escrito feito durante a entrevista.

À exceção deste caso, as transcrições foram feitas por meio de ferramenta de digitação por voz disponível no software MS Word 365. As gravações de áudio ou áudio e vídeo foram ditadas ao software pela pesquisadora.

Feitas as transcrições das entrevistas, iniciou-se a realização da sequência de análise de conteúdo proposta por Bardin (2011), consistindo de pré-análise, exploração do material e tratamento do material. Na realização da etapa de *pré-análise* com leitura das transcrições para observação de fenômenos representativos e identificação de categorias empíricas de análise identificou-se a pertinência da utilização de três das seções do questionário também como categorias de análise, sendo elas, *Linguagem de indexação*, *Formatos de metadados padrão/Protocolos computacionais* e *Mapeamentos semânticos*. A estas três somaram-se duas categorias empíricas identificadas a partir das respostas obtidas, sendo elas, *Política de indexação institucional* e *Restrições do sistema*. Compondo assim um conjunto de cinco categorias de análise.

A realização da segunda etapa de análise, *exploração do material*, realizou-se com leitura das transcrições das entrevistas para aplicação das categorias de análise definidas *à priori* com base na análise da norma ISO 25.964 partes 1 e 2. As categorias de análise com base no referencial teórico foram identificadas a partir da análise da norma ISO 25.964 partes 1 e 2, apresentada nos capítulos 2 e 3. Foram identificadas oito categorias, referentes às seções da norma que abordam conteúdo significativo para a temática da interoperabilidade semântica e de sua implementação. Em vista da característica de haverem sido identificadas no conteúdo da norma, especificamente de suas seções, optou-se por manter como nome das categorias o próprio nome das seções da norma ISO em que se espelham.

Das oito categorias, duas foram selecionadas a partir da ISO 25.964-1, apresentada no capítulo 2 desta pesquisa, sendo elas, Equivalência entre linguagens (seção 9) e Formatos de intercâmbio (seção 17). As demais 6 categorias são provenientes da ISO 25.964-2, apresentada no capítulo 3, sendo elas, Tipos de mapeamentos (seção 7), Mapeamento de equivalência (seção 8), Equivalência exata, inexata e parcial (seção 11), Uso de mapeamentos em recuperação da informação (seção 12), Lidando com a pré-coordenação (seção 13) e Apresentação de vocabulários mapeados (seção 16).

**Quadro 9** - Comparação entre tópicos das normas ISO 25.964-1 (2011) e ISO 25.964-2 (2013) e as categorias de análise elaboradas

<b>COMPARAÇÃO ENTRE TÓPICOS DAS NORMAS ISO 25.964-1 (2011) E ISO 25.964-2 (2013) E AS CATEGORIAS ELABORADAS</b>	
<b>ISO 25.964-1 (2011)</b>	
<b>Seções Selecionadas</b>	<b>Categorias elaboradas</b>
<b>9 Equivalência entre linguagens</b>	<b>Equivalência entre linguagens</b>
9.1 Geral	
9.2 Graus de equivalência	
9.3 Problemas típicos e soluções	Problemas no estabelecimento de equivalência inter-linguagens
9.4 Representação de equivalência inter-linguagens entre termos preferidos	Equivalência inter-linguagens entre termos preferidos
9.5 Equivalência inter-linguagens entre termos não-preferidos	Equivalência inter-linguagens entre termos não-preferidos
<b>17 Formatos de intercâmbio</b>	<b>Formatos de intercâmbio</b>
	Uso de formatos de metadado
<b>ISO 25.964-2 (2013)</b>	
<b>Seções Selecionadas</b>	<b>Categorias elaboradas</b>
<b>7 Tipos de mapeamentos</b>	<b>Tipos de mapeamentos</b>
7.1 Geral	Estabelecimento de mapeamentos
<b>8 Mapeamento de equivalência</b>	<b>Mapeamento de equivalência</b>
8.1 Geral	Estabelecimento de mapeamentos de equivalência
8.2 Equivalência simples	Equivalência simples
8.3 Equivalência composta	Equivalência composta acumulativa e de intersecção
<b>11 Equivalência exata, inexata e parcial</b>	<b>Equivalência exata, inexata e parcial</b>
11.1 Geral	Estabelecimento de graus de equivalência
11.2 Equivalência exata	Equivalência exata
11.3 Equivalência inexata	Equivalência inexata
11.4 Equivalência parcial	Equivalência parcial
<b>12 Uso de mapeamentos em recuperação da informação</b>	<b>Uso de mapeamentos em recuperação da informação</b>
12.1 Geral	Estágios de utilização de mapeamentos
12.2 Mediação humana	Mapeamento “manual” ou assistido por software
12.3 Exemplos de implementações de mapeamento	
12.4 Conclusões e recomendações	
<b>13 Lidando com a pré-coordenação</b>	<b>Lidando com a pré-coordenação</b>
13.1 Geral	Mapeamentos com diferentes coordenações
13.2 Mapeamento entre um tesouro e um esquema com notações dependentes do contexto	Dificuldades no estabelecimento de mapeamentos causadas por coordenação
13.3 Mapeando classes mais complexas	
<b>16 Apresentação de vocabulários mapeados</b>	<b>Apresentação de vocabulários mapeados</b>
16.1 Geral	Formas de apresentação de mapeamentos
16.2 Apresentação de registro único	
16.3 Apresentações completas baseadas em um dos vocabulários	
16.4 <i>Crosswalks</i>	

Fonte: elaboração própria

No Quadro 9 são apresentadas as seções selecionadas das duas partes da norma ISO 25.964 com o título das seções e de seus subitens primários relacionadas com as categorias elaboradas e os aspectos representativos destas seções que se considera relevante identificar em cada uma destas categorias.

O quadro 10 reúne as 13 categorias construídas para análise dos dados coletados com as entrevistas semiestruturadas, relacionando-as às etapas da análise e com a base na qual se originam.

**Quadro 10** - Sistematização da relação entre etapas da análise de conteúdo e categorias de análise

<b>Etapa da Análise</b>	<b>Categorias</b>	<b>Proveniência</b>
<b>Pré-Análise</b>	Linguagem de indexação	Protocolo de entrevista
	Formatos de metadados padrão/Protocolos computacionais	Protocolo de entrevista
	Mapeamentos semânticos	Protocolo de entrevista
	Política de indexação institucional	<i>À posteriori</i> – a partir das entrevistas
	Restrições do sistema	<i>À posteriori</i> – a partir das entrevistas
<b>Exploração do material</b>	Equivalência entre linguagens	ISO 25.964-1 seção 9
	Formatos de intercâmbio	ISO 25.964-1 seção 17
	Tipos de mapeamentos	ISO 25.964-2, seção 7
	Mapeamento de equivalência	ISO 25.964-2, seção 8
	Equivalência exata, inexata e parcial	ISO 25.964-2, seção 11
	Uso de mapeamentos em recuperação da informação	ISO 25.964-2, seção 12
	Lidando com a pré-coordenação	ISO 25.964-2, seção 13
	Apresentação de vocabulários mapeados	ISO 25.964-2, seção 16

Fonte: elaboração própria.

A realização da terceira etapa de análise bardiiana, *tratamento do material*, consiste na análise em si, com a síntese das elaborações e inferências feitas a partir do contato com o material.

#### 4.1.2 Universo de pesquisa

Quanto às instituições selecionadas para coletas de dados, estão as bibliotecas das instituições: Faculdade de Ciências e Letras de Araraquara – Unesp, Universidade Federal de São Paulo - Unifesp campus Guarulhos e Instituto de Tecnologia da Aeronáutica – ITA. Tais instituições foram identificadas no estudo de Fujita et al. (2019) como sendo instituições em que há projeto de interoperabilidade de linguagem de indexação.

A seguir apresenta-se uma caracterização das bibliotecas selecionadas para coleta de dados com entrevistas.

A Biblioteca da Faculdade de Ciências e Letras de Araraquara (FCLAr) da Unesp campus de Araraquara data o início de suas atividades em 1959, na antiga Faculdade de Filosofia, Ciências e Letras, criada como Instituto Isolado Superior do Estado de São Paulo que, posteriormente passou a integrar a Universidade Estadual Paulista “Júlio de Mesquita Filho” quando de sua criação em 1976. Este processo também suscitou o desmembramento da Faculdade em duas unidades universitárias, o Instituto de Química e o Instituto de Letras, Ciências Sociais e Educação renomeado como Faculdade de Ciências e Letras em 1989.

Assim como as demais bibliotecas da Rede de Bibliotecas da Unesp, a biblioteca da FCLAr vincula-se administrativamente à direção da unidade e tecnicamente à Coordenadoria Geral de Bibliotecas (CGB) da Unesp. A equipe da Biblioteca conta com sete bibliotecários, um coordenador e os demais divididos entre as ações da Seção Técnica de Aquisição e Tratamento da Informação (STATI) e da Seção Técnica de Referência, Atendimento ao Usuário e Documentação (STRAUD); um assistente de informática e 10 assistentes de suporte acadêmico. A biblioteca disponibiliza os serviços de disponibiliza os serviços de acesso ao Acervo digital da Unesp, Biblioteca digital de Teses e Dissertações – C@tedra; Biblioteca digital de Trabalhos de Conclusão de Curso – C@pelo; Repositório Institucional da UNESP; o catálogo de busca integrada; além de prover acesso ao Portal de Periódicos CAPES.

O acervo compõe-se de coleção, estimada em mais de 209 mil materiais entre livros, mapas, materiais audiovisuais, periódicos, trabalhos acadêmicos, entre outros; principalmente nas áreas de Humanidades, especificamente Ciências Humanas, Ciências Sociais Aplicadas, Linguística, Letras e Artes, atendendo aos cinco cursos de graduação (Administração Pública, Ciências Econômicas, Ciências Sociais, Letras e Pedagogia) e sete programas de pós-graduação (Ciências Sociais, Economia, Educação Escolar, Educação Sexual, Estudos Literários, Linguística e Língua Portuguesa e ProfLetras) oferecidos na unidade.

Quanto ao uso de linguagens de indexação, a biblioteca indicou ao estudo feito por Fujita et al. (2019) o uso do Tesouro Unesp, não sendo permitido o uso de termos advindos de outras linguagens nem o uso de linguagem natural. Para a interoperabilidade semântica, a biblioteca indicou utilizar os recursos do software Aleph para incorporar termos advindos das linguagens da Terminologia da Biblioteca Nacional (TBN), *Library of Congress Subject Headings* (LCSH), *Medical Subject Headings* (MeSH).

A Biblioteca da Escola de Filosofia, Letras e Ciências Humanas (EFLCH) foi constituída com a criação da unidade universitária do Campus Guarulhos da Universidade

Federal de São Paulo (UNIFESP). Oferece um acervo de 43 mil livros, 3.835 publicações periódicas, 270 multimeios e conta com acesso ao Portal de Periódicos Capes e à 13 bases de dados de livros digitais entre outros recursos. O acervo tem como áreas de assunto principais Ciências Humanas, Artes, Educação, Letras e Filosofia e atende aos seis cursos de graduação (Filosofia, História, Ciências Sociais, História da Arte, Letras e Pedagogia) e sete programas de pós-graduação (Educação e Saúde na Infância e Adolescência, Educação, Filosofia, Ciências Sociais, História, História da Arte e Letras) que são oferecidos na unidade universitária. A equipe da biblioteca conta com um coordenador e seis bibliotecários, além de, três assistentes em administração e 10 estagiários.

A biblioteca da EFLCH está subordinada à CRBU - Coordenadoria da Rede de Bibliotecas da Unifesp (CRBU), que disponibiliza o acesso ao acervo pelo catálogo de Busca Integrada Unifesp, com o uso do sistema de gestão de banco de dados bibliográficos Pergamum, e o Repositório Institucional UNIFESP.

Sobre a situação das linguagens de indexação na biblioteca, em estudo conduzido por Fujita et al. (2019) a biblioteca indicou que possui documento que contém a política de indexação da instituição contemplando a linguagem de indexação. Indicou também utilizar termos advindos das linguagens Terminologia da Biblioteca Nacional (TBN), *Library of Congress Subject Headings* (LCSH), Descritores em Ciências da Saúde (DeCS), Vocabulário USP, bem como o uso de termos próprios em linguagem natural. A biblioteca também indicou uso dos recursos de controle de autoridade de assunto do sistema Pergamum para realizar a interoperabilidade semântica.

A Biblioteca da Divisão de Informação e Documentação (IA-DOC) do Instituto Tecnológico de Aeronáutica (ITA) foi criada em 1950, conta com acervo na área de Ciência e Tecnologia, especialmente do setor aeroespacial, e faz parte da Rede de Bibliotecas do DCTA (Departamento de Ciência e Tecnologia Aeroespacial). A biblioteca atende aos cursos de graduação em Engenharia, que possuem seis habilitações Engenharia Aeronáutica, Engenharia Aeroespacial, Engenharia Eletrônica, Engenharia Mecânica-Aeronáutica, Engenharia Civil-Aeronáutica e Engenharia de Computação e aos cinco programas de pós-graduação com cursos de mestrado e doutorado nas áreas de Ciências e Tecnologias Espaciais, Engenharia Aeronáutica e Mecânica, Engenharia Eletrônica e Computação, Engenharia de Infraestrutura Aeronáutica e Física.

São disponibilizados pela biblioteca o catálogo de busca geral, com o uso do sistema de gestão de banco de dados bibliográficos Sophia; a Pesquisa Integrada Discovery Service do ITA; a Biblioteca Digital do ITA, base de dados de trabalhos acadêmicos de graduação,

mestrado acadêmico, mestrado profissional e doutorado; acesso à bases de dados e ao Portal de Periódicos Capes e às publicações oficiais do Comando da Aeronáutica (COMAER).

Em relação ao uso de linguagens de indexação, ao estudo feito por Fujita et al. (2019) a biblioteca indicou a existência de documento que oriente quanto à política de indexação que cobre as decisões sobre as linguagens de indexação. É admitido para a atribuição de termos o uso das linguagens NASA Thesaurus, ESJ Thesaurus, INIS Thesaurus, NCC –Thesaurus of Computing, Macrotesauro de transportes, Tesouro Spines, Inspec Thesaurus, Macrothesaurus da OECD, Thesaurus Eric Descriptors e Thesaurus do Prodasen, não sendo admitido o uso de termos em linguagem natural. A biblioteca indicou que utiliza o sistema Sophia para realizar a interoperabilidade semântica em sua linguagem própria.

#### **4.2 O polo morfológico**

O polo morfológico tem função de exposição e é o lugar de objetivação da problemática, de modo a situá-la e torná-la coerente, onde ocorre a organização dos fenômenos e da expressão da resposta à esta problemática de pesquisa. “Ao mesmo tempo, ele é o quadro operatório, prático, da representação, da elaboração, da estruturação dos objetos científicos” (BRUYNE; HERMAN; SCHOUTHEETE, p. 159).

Quanto ao polo morfológico se materializará por meio da análise em categorias dos resultados gerados pela coleta de dados e contempla a elaboração de recomendações para a construção e manutenção de linguagens de indexação com interoperabilidade semântica, construídas com base no referencial teórico instituído a partir da revisão e sistematização da literatura, referente ao polo teórico, e a partir dos dados coletados e analisados em categorias, referente ao polo técnico.

Diante do exposto com relação aos polos da pesquisa descritos, pode-se concluir de forma sintética que no método quadripolar “o polo epistemológico viabiliza os processos discursivos, o polo teórico articula os quadros de referência, o polo técnico traz os meios procedimentais de investigação e o polo morfológico exhibe os quadros de análise [...]” como elucidou Oliveira (2014, p. 27).

## **5 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS**

Nesta seção apresentam-se os resultados encontrados com o método de quadripolar de pesquisa que estrutura a pesquisa em quatro polos: polo epistemológico, polo teórico, polo técnico e polo morfológico.

Considerando os direcionamentos deste método, o polo epistemológico permitiu a identificação do objeto de pesquisa, a interoperabilidade semântica entre linguagens de indexação e a problemática da existência de diferentes linguagens de indexação em bibliotecas universitárias provenientes de diferentes origens, construídas por outras instituições ou elaboradas para uso próprio e a necessidade de integração destas em sistemas de recuperação de informação.

No polo teórico, realizou-se uma revisão de literatura de elementos teóricos sobre a construção e a manutenção de linguagens de indexação na perspectiva normativa, bem como sobre a interoperabilidade semântica. E definiu-se que para que bibliotecas universitárias possam construir e manter atualizadas linguagens de indexação que atendam às normativas internacionais necessárias à integração dos diferentes sistemas de representação de informação a interoperabilidade semântica tem importância estratégica, sendo esta a hipótese da pesquisa.

Esta revisão de literatura permitiu perceber que os procedimentos de construção e manutenção de linguagens de indexação são complexos e remontam há uma longa tradição teórica, metodológica e normativa de elaboração destes instrumentos, cuja trajetória evoluiu até os mais recentes desenvolvimentos relativos aos formatos de metadados, modelos conceituais e protocolos de troca de informações, sendo contemplados na norma internacional mais recente, a ISO 25.964. A norma ISO 25.964, parte 2, contempla a interoperabilidade semântica e métodos para alcançá-la por meio dos mapeamentos semânticos, que visam identificar relacionamentos entre os conceitos presentes em diferentes linguagens e conectá-los para permitir a realização de buscas integradas. Para colaborar nestes processos, a norma indica a utilização de formatos padrão, como o formato SKOS, para a publicação dos dados das linguagens de indexação.

Quanto ao polo técnico, estabeleceram-se os procedimentos empíricos de pesquisa que foram realizados nesta investigação: a análise de conteúdo da norma ISO 25.964, partes 1 e 2; a elaboração de um roteiro de entrevista e a realização das entrevistas nas bibliotecas indicadas anteriormente.

### **5.1 Análise das entrevistas com os procedimentos de Análise de Conteúdo**

Durante a análise dos dados coletados por meio da aplicação de entrevistas semiestruturadas foi possível levantar algumas considerações que serão apresentadas a seguir. A apresentação destes resultados será feita de acordo com as categorias elaboradas, e que foram expostas na seção anterior, e observando-se as etapas da Análise de Conteúdo de Bardin (2011).

### ***Primeira etapa - pré-análise***

A primeira etapa da análise de conteúdo de Bardin (2011) refere-se à realização de uma leitura flutuante do corpus de análise, buscando identificar fenômenos relevantes aos interesses da pesquisa. Nesta etapa, realizou-se, portanto, uma leitura preliminar das transcrições das entrevistas, buscando identificar fenômenos relevantes e recorrentes nas falas dos entrevistados. Esta leitura resultou na constatação da relevância do uso das seções do protocolo de entrevista como categorias de análise, bem como, da necessidade de elaboração de mais duas categorias empíricas, totalizando cinco categorias de análise.

### **Categoria 1 - Linguagem de indexação**

Objetivo: Verificar o estado do desenvolvimento do projeto de interoperabilidade existente na biblioteca, conforme atendimento dos critérios de seleção das bibliotecas participantes e identificar as características principais da linguagem de indexação.

Trechos:

*“Na verdade, a biblioteca separadamente não, né, a gente como faz parte da rede de bibliotecas Unesp então esse projeto é desenvolvido de forma coletiva” (Biblioteca A, Sujeito 1).*

*“[...] a gente trabalha com a linguagem integrada dentro do Aleph, que é o software que a gente usa para fazer catalogação” (Biblioteca A, Sujeito 1).*

*“Sim, tivemos o Thesaurus Aeroespacial. Na realidade esse tesouro ele está dentro do Sophia, porque os termos todos padronizados, infelizmente não no formato de um tesouro, porque ele tinha um formato de tesouro, tinha hierarquias e a gente perdeu esse arquivo [...]. E nós não julgamos como urgente a construção de um outro tesouro como era aquele porque eu tenho tudo dentro do Sophia, estou com o meu vocabulário controlado todo lá. Então o termo que entra eu já padronizo, já verifico, o termo para tá lá ele não é um termo*



*sugerido, ele já é um termo aceito. Então o processo que a gente utilizou até para a construção do Thesaurus Aeroespacial ela mantém essa filosofia, mantenho até hoje” (Biblioteca B, Sujeito 2).*

*“Sim, existe um projeto [para a construção de uma linguagem de indexação], mas está mais no papel, seria um projeto mais individual do bibliotecário-chefe” (Biblioteca C, Sujeito 4).*

*“O coordenador da biblioteca tem interesse em criar um vocabulário nosso, de Humanas para o campus, ele gostaria de começar a pensar isso ainda esse ano. Há uma bibliotecária nova que entrou no ano passado para trabalhar no setor de referência, que estava precisando desenvolver, estávamos nos preocupando com outras questões, e ela detectou a falta que o vocabulário faz.” (Biblioteca C, Sujeito 3).*

Inferências: Das três bibliotecas questionadas, uma possui um projeto de desenvolvimento de linguagem de indexação própria em conjunto com a rede de bibliotecas da universidade a qual faz parte, uma possui uma linguagem própria já estabelecida e em uso, e a terceira declara haver um projeto não iniciado para a construção de uma linguagem própria. Percebe-se nesta categoria que os sujeitos entrevistados, nem sempre reconhecem que os dados de autoridade de assunto presentes em seu sistema de gerenciamento constituem uma linguagem de indexação, mesmo que tenha sido composto seguindo critérios para tal. Isso está perceptível nas respostas das bibliotecas B e C, especialmente percebe-se na biblioteca B a falta de uma opção de visualização sistemática da linguagem, tal como os tesouros em geral oferecem, para compreenderem aqueles dados como efetivamente uma linguagem de indexação.

## **Categoria 2 - Formatos de metadados padrão/Protocolos computacionais**

Objetivo: Identificação dos formatos e padrões de metadados utilizados visando o estabelecimento de interoperabilidade semântica entre linguagens de indexação.

Trechos:

*“Ah, a gente usa o mesmo formato de metadados que a gente usa para catalogação dos livros, mas claro usando os campos específicos do MARC no que diz respeito a catalogação de assunto ou de autor. [...] É, MARC Autoridades.” (Biblioteca A, Sujeito 1).*

*“Às vezes, por exemplo, uma das coisas que a gente achou bem assim diferente, foi quando foi criar os cabeçalhos para personagens fictícios, quando você usa o nome da pessoa seguido do personagem fictício, por exemplo, Robson Crusóé aí entre parênteses personagem fictício. A BN tratava ele num campo e a LC em outro completamente diferente, um tratava como assunto tópico e a outra tratava como nome pessoal. E aí era uma diferença que no começo a gente ficou “Como que a gente faz?” e aí quem acabou decidindo foi o grupo né, que trata de indexação, e aí fizeram reuniões... e aí decidiram como é, para ficar padronizado né, por que às vezes a pessoa, uma de nós da rede importava da BN e ficava num determinado campo e quem importava da LC ficava em outro e era a mesma coisa. [...] O entendimento que tem de se aquela pessoa, o fictício, ele é só um assunto tópico ou se, apesar de ser fictício, ele é uma pessoa sim e deveria ser tratado como um campo 100” (Biblioteca A, Sujeito 1)*

*“Os sistemas hoje eles têm campo para tudo, tem MARC, tem tudo. Antigamente não era assim. O que a gente tinha era uma norma ISO que diz dado em rede de acervo, de informação bibliográfica, tem que ser passado dessa maneira, que é o que permite que os bancos se juntem e a gente consiga pesquisar” (Biblioteca B, Sujeito 2).*

*“A biblioteca, eu não sei se eu comentei com você, mas ela tem um analista de sistemas aqui dentro. Então ela é diferenciada por isso também, ela não desenvolve os nossos programas, ela já desenvolveu, tanto que a biblioteca digital do ITA ela é quem criou, hoje ela não consegue mais atualizar, então por isso que a gente voltou para o Sophia, está voltando para o Sophia. Porque todos os dados da minha biblioteca digital, o que que ela faz, ela pega os metadados daqui e cria uma máscara aqui para consultar, nada mais justo que fique tudo dentro do Sophia, porque ele já tem essa possibilidade” (Biblioteca B, Sujeito 2)*

*“MARC 21 Autoridades” (Biblioteca C, Sujeito 3).*

Inferências: Todas as três bibliotecas indicaram a utilização do formato de metadados MARC 21 Autoridades para a representação dos dados de autoridade. A norma ISO 25.964-1 reconhece o MARC21 Autoridades como um formato de metadados para dados de autoridade extensamente utilizado para o registro de linguagens de indexação por bibliotecas, porém a norma também reconhece que este formato não é adequado caso a linguagem seja construída com o uso do modelo de dados que a norma descreve, recomendando para tal situação o uso do formato *XML schema* que a própria norma traz.

### **Categoria 3 - Mapeamentos semânticos**

Objetivo: Identificar o estágio de desenvolvimento do projeto de interoperabilidade existente na biblioteca e a compreensão dos respondentes acerca das dificuldades encontradas neste processo.

Trechos:

*“A questão é que, na hora de fazer os cabeçalhos, na hora de pensar nessa construção, a gente acaba usando essas que são pós-coordenadas, que é a LC que é BN, que você tem essa liberdade para construir os cabeçalhos” (Biblioteca A, Sujeito 1).*

*“Quando a gente precisa fazer buscas externas, sejam na LC ou na BN, a gente consegue encontrar mais os termos que a gente precisa, pois são termos assim mais consolidados, e às vezes, são bem específicos [...] são termos mais consolidados. São raras as exceções, eu acho que até hoje a gente precisou uma única vez solicitar a criação de um termo que não tinha nem na LC nem na BN e aí a gente passou por todo aquele processo de preencher um formulário, de pedir ajuda ao docente específico da área, de levantar fontes, mas foi uma única vez, nas outras vezes que a gente precisou de assuntos eles estavam contemplados ou na BN ou na própria LC, se já não estavam na própria linguagem da UNESP” (Biblioteca A, Sujeito 1).*

*“E é claro como a gente usa de linguagens diferentes, inclusive, além do que eu, aqui a gente só usa a BN e a LC, mas eu vejo nas instruções que vêm, por exemplo, o pessoal de saúde usa outras linguagens específicas, então é mais só a gente aparar essas pequenas arestas, mas acaba se tornando uma linguagem mais completa possível, pelo menos no meu ponto de vista” (Biblioteca A, Sujeito 1).*

*“Não, para os termos de assuntos só pós-coordenadas, pré-coordenação apenas na classificação com a CDU. Usamos tesouros, o Tesouro da NASA, a LCSH já usei mais no passado” (Biblioteca B, Sujeito 2).*

*“Não, o ITA ele é muito específico na área do conhecimento, eu tenho Engenharia Aeronáutica, Engenharia Mecânica e Aeronáutica, Engenharia Aeroespacial, Engenharia da Computação, Engenharia Civil, Engenharia Eletrônica, então é muito específico, é muito específico” (Biblioteca B, Sujeito 2).*

*“Eu tenho algumas posturas o que eu estou percebendo que nem todo mundo compartilha dela, eu acho que tem que ser um banco de dados único, já chega você ter outras informações fragmentadas, por isso que a gente tem tudo. Já fui vaiada porque repositório institucional é repositório institucional, eu falei o meu chama biblioteca da produção [...]” (Biblioteca B, Sujeito 2).*

*“Foi feita uma reunião coletiva das bibliotecas para discutir sobre uso de linguagem, qual linguagem ou linguagens usar para não ficar com o catálogo desorganizado. Foi decidido que a primeira linguagem seria a BN. Em Guarulhos usamos exclusivamente a BN, em SP e Diadema, campus com ênfase nas áreas de saúde, utilizam o DeCS (para termos da área saúde) e a BN (para termos mais genéricos), foi necessário definir isso em reunião, porque já houve caso de o bibliotecário usar o termo do DeCS de áreas mais genéricas e o conflitar com a BN. Temos problemas de comunicação entre as unidades o que prejudica a realização de adaptações ou mudanças nos vocabulários. Para a área de História da Arte uma professora indicou vocabulários do MASP e Itaú cultural. Também chegamos a utilizar um meta-buscador de vocabulários [BARTOC]<sup>9</sup>” (Biblioteca C, Sujeito 4).*

*“Quando o termo está pré-coordenado na BN é mantido, especialmente por causa das remissivas, não é favorável “é uma faca de dois gumes”.*

---

<sup>9</sup> Catálogo *Basel Register of Thesauri, Ontologies & Classifications* (BARTOC, disponível em: <https://bartoc.org/>)

*Mas, se precisar importar da BN um termo pré-coordenado ele será mantido” (Biblioteca C, Sujeito 4).*

*“Vemos muita dificuldade em incorporar outras linguagens, acabando sendo atendidos pela BN, em geral. Por mais críticas à BN/LCSH, é um vocabulário utilizado no mundo todo, por exemplo, na OCLC já tem algumas indicações de termos e na prática nos atende. Mesmo vendo algumas lacunas, evitamos utilizar outros vocabulários, pelo receio de encavalhar termos de vocabulários diferentes. Acabamos mais localizando termos nas outras linguagens, como no Itaú Cultural ou no MASP, e depois buscamos um correspondente na BN para tentar manter uma consistência. E em último caso criamos o termo.” (Biblioteca C, Sujeito 4).*

Inferências: Quanto ao uso de linguagens com coordenação diferentes, duas bibliotecas afirmaram que utilizam apenas linguagens pós-coordenadas como fonte para o processo de compatibilização, indicando o uso das linguagens LCSH, TBN e Tesouro da NASA. Já a terceira instituição indicou que procuram utilizar exclusivamente a TBN ou LCSH. Quanto ao uso de linguagens com níveis de especificidade diferentes, a biblioteca A biblioteca indicou que a utilização das linguagens TBN e LCSH supre as necessidades de especificidade das áreas de assunto atendidas pela biblioteca e a biblioteca B indicou que o uso de linguagens muito específicas é predominante, por atenderem áreas do conhecimento com evolução mais rápida. Já a biblioteca C indicou que embora a necessidade por termos mais específicos exista, dá preferência ao uso apenas da TBN, novamente houve demonstração de dificuldade de compatibilizar linguagens diferentes. Os respondentes indicaram que utilizam duas linguagens específicas, linguagens das instituições culturais Itaú Cultural e MASP (Museu de Arte de São Paulo), apenas como fontes externas para posteriormente localizar um termo correspondente na TBN. A biblioteca C também foi a única que relatou outras dificuldades no processo de gestão da linguagem, indicou que há dificuldades de organização entre as diversas bibliotecas da rede, que atuam em áreas de assunto diferentes, e que estas dificuldades têm sido percebidas como impedimentos para as ações necessárias para a atualização da linguagem.

#### **Categoria 4 - Política de indexação institucional**

Objetivo: Identificar elementos relativos a discussões institucionais sobre os procedimentos de indexação e gestão das linguagens de indexação.

Trechos:

*“Agora se é algum caso como foi esse do personagem, além de estabelecer como isso vai ser usado, a gente sempre recebe um feedback dizendo o porquê, ah a decisão foi essa então a gente faz assim e aí eu, por exemplo, quando eu recebo estagiárias todas essas informações que a gente têm, desde as que estão formalmente lá no site quanto essas que vão sendo trocadas por e-mail, que vão sendo respostas das nossas dúvidas, a gente vai passando, para deixar todo mundo com o mesmo nível de conhecimento do que tem que fazer no sistema” (Biblioteca A, Sujeito 1).*

*“O processamento técnico, ele já faz isso já é fluxo internalizado. [...] A mesma coisa com os números da CDU, Mecânica do voo é um número que não existe, o ITA criou um número para mecânica do voo, por que esse é um assunto importante e aquele acervo precisa estar separado, o meu usuário, por que a biblioteca é de livre acesso, tem que entrar no corretor da Mecânica do voo e se achar, ele não achou aquele autor, mas ele tem 500 para ela procurar, então a gente criou. [...] Ele tem um processo, tudo que a gente faz tem uma cadeia de comando, tem um processo. É um número novo tenho que criar, ele é criado a partir e é anotado. É um índice novo, é um termo novo, ele segue o processo. Ele tem um fluxo” (Biblioteca B, Sujeito 2).*

*“Foi feita uma reunião coletiva das bibliotecas para discutir sobre uso de linguagem, qual linguagem ou linguagens usar para não ficar com o catálogo desorganizado. Foi decidido que a primeira linguagem seria a BN. Em Guarulhos usamos exclusivamente a BN, em SP e Diadema, campus com ênfase nas áreas de saúde, utilizam o DeCS (para termos da área saúde) e a BN (para termos mais genéricos), foi necessário definir isso em reunião, porque já houve caso de o bibliotecário usar o termo do DeCS de áreas mais genéricas e o conflitar com a BN. Temos problemas de comunicação entre as unidades o que prejudica a realização de adaptações ou mudanças nos vocabulários” (Biblioteca C, Sujeito 4).*

Inferências: Identifica-se nesta categoria que duas instituições, as bibliotecas A e B, relataram situações em que a existência de uma política de tratamento da informação, uma política de indexação ou mais generalizada, têm papel relevante. O respondente da Biblioteca A colocou uma situação de uso de campos específicos do formato de metadados para dados de autoridade que surgiu com a indexação de obras de ficção (tratar personagens fictícios como assunto ou como pessoa) e que foi resolvida e amplamente divulgada à equipe de trabalho, de modo que não exista inconsistências na representação destes dados. Na Biblioteca B, o respondente relatou a situação em que a política está internalizada pela equipe de trabalho, fato que ocorre com o tempo de trabalho efetuando os processos com orientações específicas. Já a situação exposta pelos respondentes da Biblioteca C referem-se à necessidade de negociação de procedimentos de compatibilização das linguagens para solucionar casos de linguagens com níveis de especificidade diferentes. Todas estas situações, entre outras, podem surgir durante o processo de construção e manutenção de linguagens de indexação e esta negociação e divulgação das decisões tomadas é essencial para a consistência dos processos realizados, principalmente tratando-se de instituições descentralizadas, como o caso das bibliotecas A e C. Neste sentido, a norma ISO 25.964-1 orienta em sua seção 13 (Gerenciando a Construção e Manutenção de Tesouros) para a determinação de objetivos e o estabelecimento de responsabilidades para a equipe responsável pelo trabalho com a linguagem de indexação.

### **Categoria 5 - Restrições do sistema**

Objetivo: Identificar situações relacionadas à capacitação da equipe responsável pelo projeto de linguagem, os recursos em geral à disposição, tais como a infraestrutura da instituição, recursos humanos, entre outros elementos relacionados.

Trechos:

*“É tudo bem na raça, tivemos que encarar o desafio do processamento técnico aqui, bem por conta de termos uma equipe pequena e grande quantidade de itens a ser catalogados, junto com as cobranças de resultados, ficamos muito limitados a resolver os problemas primeiro para depois se preocupar com outras questões” (Biblioteca C, Sujeito 3).*

Inferências: Nesta categoria, identificam-se dificuldades para a realização das atividades de gestão da linguagem, especialmente por questões de restrição de recursos humanos diante da

complexidade e quantidade do trabalho a ser realizado que se soma às demais atividades sob a responsabilidade dos sujeitos.

#### *Síntese da etapa de pré-análise:*

Como esperado, todas as bibliotecas indicaram utilizar o formato MARC21 Autoridades como formato de representação dos metadados das linguagens de indexação. Este formato é amplamente utilizado em sistemas de gerenciamento de bibliotecas, bem como, em sistemas de construção de linguagens de indexação. Porém, como mencionado anteriormente, caso a linguagem seja construída com base no modelo de dados sugerido pela norma ISO 25.964-2 o formato MARC21 não será adequado. Softwares de construção de linguagens de indexação, como o já mencionado TemaTres, oferecem a possibilidade de trabalhar com outros formatos de metadados como o Skos-Core (*Simple Knowledge Organization System*), MADS (*Metadata Authority Description Schema*) entre outros.

Foram identificadas dificuldades para compatibilização de diferentes linguagens de indexação, compreende-se que a realização deste procedimento depende de uma análise conceitual minuciosa e, antes disso, é preciso distinguir entre o termo e o conceito que este representa na linguagem.

Em síntese, nesta categoria identificamos que a Biblioteca C, que indicou que seu projeto de construção de uma linguagem de indexação não está efetivamente sendo desenvolvido, foi aquela que relatou mais dificuldades na realização dos processos. Compreende-se que estas dificuldades decorrem da ausência de definição do projeto com relação aos seus objetivos, as atividades a serem desenvolvidas e a divisão de responsabilidades entre a equipe envolvida, causando incertezas mesmo na realização das atividades cotidianas relativas à linguagem de indexação.

#### ***Segunda etapa - exploração do material***

A segunda etapa da análise de conteúdo de Bardin (2011) refere-se à realização da exploração do material. Nesta etapa, trabalhou-se com as categorias definidas *à priori*, com base na análise realizada no conteúdo das normas ISO 25.964-1 e ISO 25.964-2. Foram elaboradas oito categorias para esta etapa de análise. Porém em cinco destas categorias não houve identificação de ocorrências significativas nas falas dos entrevistados, sendo elas: Categoria 3 - Tipos de mapeamentos, Categoria 4 - Mapeamento de equivalência, Categoria 5 - Equivalência exata, inexata e parcial, Categoria 6 - Uso de mapeamentos em recuperação da



informação e Categoria 8 - Apresentação de vocabulários mapeados. Dessa forma apresenta-se apenas a análise das categorias em que foi possível identificar ocorrências.

### **Categoria 1 - Equivalência entre linguagens**

Objetivo: Identificar se estão desenvolvendo mapeamentos de equivalência inter-linguagens e se foi observada a existência de problemas no estabelecimento de mapeamentos de equivalência.

Trechos:

*“Acabamos mais localizando termos nas outras linguagens, como no Itaú Cultural ou no MASP, e depois buscamos um correspondente na BN para tentar manter uma consistência. E em último caso criamos o termo” (Biblioteca C, Sujeito 4).*

Inferências: Percebe-se nesta categoria uma situação de uso de uma ação inicial de mapeamento de equivalência entre conceitos presentes em diferentes linguagens, porém, sendo um uso bastante “rústico” deste procedimento. A biblioteca utiliza essa ação apenas para localizar o termo correspondente àquele que existe nas linguagens especializadas, sem estabelecer efetivamente o mapeamento, que aconteceria com a declaração de mapeamento e a indicação da linguagem alvo.

### **Categoria 2 - Formatos de intercâmbio**

Objetivo: Identificar se houve definição quanto ao formato de metadados utilizado para o desenvolvimento do projeto de desenvolvimento/manutenção de linguagens de indexação visando o alcance da interoperabilidade semântica ou seguindo quais parâmetros.

Trechos:

*“[...] a gente trabalha como a linguagem integrada dentro do Aleph, o que é o software que a gente usa para fazer catalogação. [...] É, MARC Autoridades” (Biblioteca A, Sujeito 1).*

*“Exatamente. O entendimento que tem de se aquela pessoa, o fictício, ele é só um assunto tópico ou se, apesar de ser fictício, ele é uma pessoa sim e deveria ser tratado como um campo 100” (Biblioteca A, Sujeito 1).*

*“Os sistemas hoje eles têm campo para tudo, tem MARC, tem tudo. Antigamente não era assim. O que a gente tinha era uma norma ISO que diz dado em rede de acervo, de informação bibliográfica, tem que ser passado dessa maneira, que é o que permite que os bancos se juntem e a gente consiga pesquisar” (Biblioteca B, Sujeito 2).*

*“Na realidade esse tesouro ele está dentro do Sophia, porque os termos todos padronizados, infelizmente não no formato de um tesouro, por que ele tinha um formato de tesouro, tinha hierarquias e a gente perdeu esse arquivo porque ele era feito ainda no IBM, o IBM gera aquelas fitas, no IBM antigo né, você gerava aquelas fitas e uma dessas fitas se perdeu e se perdeu o equipamento. E nós não julgamos como urgente a construção de um outro tesouro como era aquele porque eu tenho tudo dentro do Sophia, estou com o meu vocabulário controlado todo lá” (Biblioteca B, Sujeito 2).*

*“Por muito tempo, desde a implantação das bibliotecas o software utilizado era o PHL, em 2015 foi adquirido o software Pergamum. No PHL o controle de vocabulário era bem ruim, com a migração para o Pergamum os problemas com o controle de vocabulário ficaram mais evidentes. E estamos ainda corrigindo esses problemas que ficaram. O Pergamum atende bem, em comparação com o PHL é muito melhor, as dificuldades que aparecem são por conta dos problemas que ficaram no vocabulário” (Biblioteca C, Sujeito 3).*

*“No PHL o campo de assunto não tinha relação com a base de autoridades, então não conseguia colocar remissivas, era um campo solto. E quando migramos para o Pergamum o campo de assunto veio para um campo de assunto local, 697. E agora o Pergamum é muito mais rico os campos 600 conversam com a base de autoridades. E como somos só nós dois para fazer isso, acabamos, conforme a catalogação na busca do termo para indexar, quando percebemos problemas nos termos vamos limpando” (Biblioteca C, Sujeito 4).*

*“Na migração grande parte do acervo que estava catalogado e tinha os assuntos no PHL, os assuntos foram para esse campo 697, que hoje na busca por assunto no catálogo não recupera, só recupera se for*

*pesquisa geral, então a dificuldade hoje é mais por conta disso” (Biblioteca C, Sujeito 3).*

Inferências: Todas as três bibliotecas indicaram situações que revelam a importância dos formatos de metadados para dados de autoridade padronizados em relação ao trabalho com linguagens de indexação. A situação colocada pelos respondentes da Biblioteca B remete também à necessidade de preservação digital, o que também é preciso ser debatida quanto a questão dos formatos de intercâmbio de dados. Já a situação colocada pela Biblioteca C revela perda de informações decorrentes de mudança do sistema de gestão de acervo em decorrência de inconsistências nos campos dos formatos de metadados utilizados. O problema relatado com o primeiro software utilizado, que não realizava a interconexão entre base de dados bibliográficos e base de dados de assunto, sem essa comunicação não existe o efetivo controle de vocabulário como já relatado em Fujita, Piovezan e Santos (2017). Essa questão também denota a importância do uso de um software adequado de construção de linguagens de indexação, que possam ser integrados ao software de gestão de acervo e dessa forma permitir a intermediação entre registros bibliográficos e registros de autoridade de assunto.

### **Categoria 7 - Lidando com a pré-coordenação**

Objetivo: Identificar se o projeto de desenvolvimento/manutenção de linguagens de indexação envolve trabalho com linguagens que possuem diferenças na coordenação dos termos e se foram observadas dificuldades no estabelecimento de mapeamentos que possam ter sido causadas por estas diferenças.

Trechos:

*“Não, para os termos de assuntos só pós-coordenadas, pré-coordenação apenas na classificação com a CDU. Usamos tesouros, o Tesouro da NASA, a LCSH já usei mais no passado” (Biblioteca B, Sujeito 2).*

*“Quando o termo está pré-coordenado na BN é mantido, especialmente por causa das remissivas, não é favorável “é uma faca de dois gumes”. Mas, se precisar importar da BN um termo pré-coordenado ele será mantido” (Biblioteca C, Sujeito 4).*

Inferências: Nesta categoria foi identificada a situação presente na Biblioteca C, que procuram utilizar exclusivamente a TBN ou LCSH, principalmente por encontrarem dificuldades para a realização da compatibilização com outras linguagens e o receio de compor uma linguagem desorganizada, os respondentes reconhecem que a TBN possui termos pré-coordenados – por sua constituição derivada da LCSH, inicialmente uma linguagem pré-coordenada – o que acarreta dificuldades em decompor estes termos, fazendo com que tenham optado por mantê-los com a coordenação original.

#### *Síntese da etapa de exploração do material:*

Quanto à etapa metodológica de exploração do material, não houve ocorrências nas falas dos entrevistados com relação à cinco das oito categorias de análise elaboradas *à priori*, impedindo a realização da análise destas categorias. Esta ausência decorreu principalmente do fato de que os elementos abordados nestas categorias não eram de conhecimento dos entrevistados, não estando identificados como “parte” dos projetos de interoperabilidade nessas bibliotecas. Este desconhecimento advém do desconhecimento do próprio documento normativo e das discussões conceituais e metodológicas que nele são abordadas.

Considerando que esta segunda etapa foi desenvolvida com base na literatura, o esperado com relação às respostas dos entrevistados era que fossem abordados os tipos de mapeamentos que estariam sendo realizados (equivalência, hierárquicos e associativos), se seria dada a preferência à um tipo, como por exemplo, o mapeamento de termos preferidos (mapeamento de equivalência), ou todos; qual o tipo de apresentação das declarações de mapeamentos, assim como, quaisquer particularidades relativas a realização destas atividades em cada uma das bibliotecas.

Essas categorias não representadas nesta etapa da pesquisa referem-se de modo central aos mapeamentos entre conceitos, apresentados pela norma ISO 25.964-2 (2013) como o principal método para estabelecimento de interoperabilidade semântica entre linguagens. Isso nos leva a compreender o desconhecimento destes elementos e do conteúdo da norma em geral, como uma questão que deverá ser mais abordada em pesquisas futuras para contribuir com a interoperabilidade semântica.

## **5.2 Síntese dos resultados: recomendações para construção e manutenção de linguagens de indexação visando a interoperabilidade semântica**

Considerando a realização dos procedimentos de pesquisa aqui relatados, que abordaram desde os pressupostos teóricos sobre as linguagens de indexação, sua elaboração e manutenção

com foco na interoperabilidade semântica, até a realização do estudo empírico de identificação e análise de projetos de linguagem de indexação visando a interoperabilidade semântica, recomenda-se que projetos visando esse fim observem o seguinte:

- Atendimento à normalização internacional vigente;
- Compor uma equipe de trabalho, reunindo quadro de pessoal que possua expertise nos processos relativos à construção e gestão de linguagens de indexação, com atribuição de responsabilidades aos membros desta equipe;
- A equipe de trabalho deve ser composta por membros do maior número e diferentes unidades em casos de instituições descentralizadas;
- Definir os objetivos da linguagem de indexação e o público-alvo desta linguagem;
- Realizar diagnóstico dos recursos humanos, materiais e de vocabulário disponíveis;
- Definir o software de gestão de tesouros que será utilizado;
- Definir o modelo de mapeamento que será adotado para o estabelecimento de relações e a direção dos mapeamentos;
- Considerar a realização dos mapeamentos de acordo com o procedimento geral indicado no item 14.1 da norma ISO 25.964 parte 2;
- Estabelecer mapeamentos tendo como base análise conceitual dos conceitos envolvidos;
- Utilizar formato de metadados padronizado para descrição dos dados de autoridade que compõem a linguagem de indexação;
- Apresentar as declarações de mapeamento indicando o tipo de mapeamento entre os conceitos com o uso das notações padronizadas e o uso de identificadores únicos e persistentes, tais como o URI;
- As decisões tomadas pela equipe de trabalho devem ser divulgadas aos demais bibliotecários que farão uso da linguagem de indexação, para garantir a consistência dos processos realizados.

## 6 CONSIDERAÇÕES FINAIS

Para a realização do trabalho de pesquisa que aqui se apresenta, propôs-se a realização de um estudo teórico e metodológico sobre a interoperabilidade semântica entre linguagens de indexação no contexto da política de indexação de modo a elencar recomendações necessárias à implantação de linguagem de indexação interoperável destinada a ser compartilhada em sistemas integrados de bibliotecas universitárias.

Desse modo, apresentou-se como problema de pesquisa a disseminação de vocabulários com níveis de controle diversos construídos individualmente por bibliotecas universitárias brasileiras como fator de isolamento, por não atenderem às normativas internacionais e assim não permitirem o alcance da interoperabilidade semântica com outros vocabulários.

A hipótese colocada para esta pesquisa compreende que a interoperabilidade semântica tem importância estratégica para que bibliotecas universitárias possam construir e manter atualizadas linguagens de indexação que atendam às normativas internacionais necessárias à integração dos diferentes sistemas de representação de informação.

Agregar interoperabilidade semântica às linguagens de indexação em bibliotecas universitárias, sustenta a tese de que garantirá a melhora nos processos de gestão destas linguagens, contribui para sua atualização, incorporação de termos, bem como a percepção no aumento da qualidade da indexação realizada.

O problema de pesquisa colocado foi reafirmado durante a realização dos procedimentos empíricos, com as entrevistas semiestruturadas, no sentido em que se percebe que as três instituições abordadas desenvolvem vocabulários próprios seguindo parâmetros que pouco se inter cruzam. As entrevistas também permitiram identificar dificuldades no desenvolvimento de projetos de construção e manutenção de linguagens de indexação, tais como, compatibilizar conceitos de diferentes linguagens para compor uma linguagem nova e o estabelecimento de relações entre termos e conceitos. Tais dificuldades afetam as ações cotidianas de gestão da linguagem de indexação e, também, têm efeitos no uso desta linguagem pelos usuários.

Estas dificuldades, são abordadas pela norma ISO 25.964, que apresenta orientações sobre como desenvolver linguagens de indexação e sobre a interoperabilidade entre linguagens de indexação. Nas entrevistas realizadas revelou-se que os projetos em desenvolvimento nas instituições analisadas não contemplam diretrizes da referida norma, especialmente no que se refere à interoperabilidade semântica. Nesse sentido buscou-se elaborar um conjunto de recomendações para contribuir com o desenvolvimento de projetos tais como os analisados.

Estabeleceram-se como objetivos específicos, **1)** realizar estudo teórico e metodológico com revisão da literatura nacional e internacional sobre elaboração e manutenção de linguagens de indexação com foco na interoperabilidade semântica; **2)** analisar o processo de elaboração de linguagens de indexação sob o aspecto das normas internacionais focalizando a interoperabilidade semântica; **3)** identificar e analisar projetos de interoperabilidade semântica de linguagens de indexação, e **4)** elaborar recomendações para a construção de linguagens com interoperabilidade semântica.

Partindo destes objetivos, almejou-se alcançar o objetivo geral de oferecer subsídios teóricos e metodológicos à pesquisa sobre a interoperabilidade semântica de linguagens de indexação no contexto teórico da Organização e Representação do Conhecimento.

Para isso foi escolhido para o desenvolvimento desta investigação a abordagem com o método de quadripolar de pesquisa que estrutura a pesquisa em quatro polos: polo epistemológico, polo teórico, polo técnico e polo morfológico. Este método mostrou-se adequado a realização da pesquisa, a interconexão dos quatro polos permitiu manter uma atitude ativa e de permanente autorregulação com relação à dinâmica da pesquisa, bem como foi adequado para coordenar a utilização das diferentes técnicas de coleta e análises dos dados.

Como meio para atingir o objetivo específico **1)** foi elaborado um capítulo teórico que se dirigiu ao tema da linguagem de indexação suas características e desafios sob o enfoque da organização do conhecimento em Ciência da Informação. Neste capítulo, procuramos abordar as linguagens de indexação tendo o tesouro como o instrumento *sui generis* desta classe, que deste modo nos permitiu estudar desde o contexto do surgimento do tesouro e o percurso teórico, metodológico e normativo relativo às suas funções, características e objetivos nos processos de representação e recuperação da informação, culminando na apresentação da norma internacional ISO 25.964-1, vigente desde 2011, que aborda o processo de construção e manutenção de tesouros visando ao enfrentamento dos mais atuais desafios da organização e representação do conhecimento no âmbito digital e interconectada. Também se considerou relevante apresentar neste capítulo uma revisão sobre o processo de determinação da linguagem de indexação inserida no âmbito da política de indexação.

Para atingir o objetivo específico **2)** foi elaborado um capítulo teórico direcionado ao tema interoperabilidade semântica entre linguagens de indexação na perspectiva da organização do conhecimento em Ciência da Informação. O capítulo desenvolvido buscou discutir o conceito de interoperabilidade semântica e os objetivos de sua implementação nos ambientes de bibliotecas que se aplicam às linguagens de indexação com base na literatura recente sobre o tema no campo da organização do conhecimento em Ciência da Informação. Procurou-se

compor este capítulo com base na norma internacional ISO 25.964-2 que aborda a interoperabilidade entre linguagens de indexação e considerou-se relevante a explanação do conteúdo desta norma, ressaltando os elementos essenciais para o alcance da interoperabilidade, sendo estes os mapeamentos semânticos, esta forma, o capítulo procurou discutir os modelos de mapeamentos apresentados pela norma ISO 25.964-2. Buscou-se também discutir os formatos de metadados para a representação de dados de autoridade de assunto existentes, prevendo sua importância para a interoperabilidade semântica.

A partir dos quadros teóricos abordados que compuseram a formalização do polo teórico da pesquisa, compreende-se que os desenvolvimentos tecnológicos advindos da evolução da web fomentam uma nova maneira de lidar com a informação disponível nas bibliotecas universitárias. Estes desenvolvimentos permitem que bibliotecas elaborem sistemas que agrupem diversas bases de dados e repositórios que juntos oferecem acesso a informações em nível incalculável. Entretanto, embora formatos para intercâmbio em nível tecnológico existam, estes garantem a interoperabilidade de dados, porém não a compreensão compartilhada, objetivo da interoperabilidade semântica. Para completar estes desenvolvimentos, exige-se destes sistemas a compreensão do conteúdo trocado, que pode ser facilitada com a realização dos mapeamentos semânticos entre os conceitos de diferentes linguagens de indexação. Tais temas são complexos, porém são de capital importância para o trabalho bibliotecário no contexto atual de informação digital e interconectada.

Dessa forma, os resultados teóricos indicaram a importância de investigar o conhecimento e a aplicação das diretrizes normativas sobre construção de linguagens de indexação nas bibliotecas que possuem projetos de interoperabilidade semântica. Por isso, como meio para alcançar o objetivo específico, **3)** identificar e analisar projetos de interoperabilidade semântica de linguagens de indexação, a investigação contou em seu polo técnico com a realização de entrevistas semiestruturadas em três bibliotecas que haviam sido identificadas, com base na pesquisa realizada por Fujita et al. (2019), como tendo projeto de construção/atualização de linguagem de indexação com interoperabilidade semântica. Estas pesquisas foram analisadas em categorias de assunto com a utilização dos procedimentos advindos da Análise de Conteúdo de Bardin (2011).

Nestas análises, os elementos que mais se evidenciam são os que não foram mencionados. Isto porque os sujeitos respondentes das entrevistas não puderam falar sobre os procedimentos que foram discutidos na etapa teórica desta pesquisa, já que não são realizados nestas instituições. Percebeu-se o grande desconhecimento sobre as normas internacionais de construção de linguagens de indexação estudadas por esta investigação, mesmo a primeira parte



desta norma, ISO 25.964-1 (2011) que aborda especificamente os processos de construção de tesauro para a recuperação da informação, estando para completar 10 anos de sua publicação original já no próximo ano.

É preciso considerar que a falta de conhecimento sobre estes conteúdos se dá também por questões de acessibilidade da norma, cuja publicação está no idioma inglês e tem custo de aquisição bastante severo. Não se tem conhecimento da existência de iniciativas para possibilitar o aumento desta acessibilidade às normas ISO 25.964-1 e 2 para a comunidade bibliotecária brasileira, porém esta mostra-se como uma etapa essencial no momento para lidar com o problema do isolamento causado pela diversidade de vocabulários com níveis de controle diversos construídos individualmente por bibliotecas universitárias brasileiras. Sem o conhecimento dos parâmetros atuais para o processo de construção e manutenção de linguagens por parte da comunidade bibliotecária brasileira o problema da difusão de vocabulários e bases de autoridade de assunto persistirá.

Com relação ao nosso último objetivo específico, **4)** elaborar recomendações para a construção de linguagens com interoperabilidade semântica, foi possível a identificação por meio da análise da norma ISO 25.964 partes 1 e 2, dos elementos constituintes de um projeto de interoperabilidade semântica, seguindo-se a descrição e debate sobre os elementos que constituem estas normas e, por fim, como síntese dos resultados encontrados ao longo da pesquisa o item 5.2 apresentou um conjunto inicial de recomendações de práticas para o desenvolvimento destes projetos. Espera-se que esse construto funcione como um passo inicial para o desenvolvimento de projetos deste gênero por parte de bibliotecas universitárias brasileiras.

Identificou-se por meio do contato com as bibliotecas pesquisadas que o domínio profissional adota o uso do termo “linguagem”, seja acompanhado dos qualificadores “documentária” ou “de indexação”, para se referir à classe de instrumentos voltados a representação por assunto. Isto reforça nossa opção pelo uso do termo “linguagem de indexação” que além de estar próximo da comunidade de uso também evidencia a atividade de indexação que é realizada pelo bibliotecário catalogador.

Uma parcela dos sujeitos pesquisados, bibliotecários universitários, mostraram-se presos ao princípio de não alterar os instrumentos de trabalho, nisto incluindo a linguagem de indexação, mesmo em situações em que a linguagem não os atende, desta forma não consideram a opção de uso dos mapeamentos como forma de aperfeiçoamento destas linguagens.

Conclui-se que o desconhecimento dos mapeamentos semânticos, enquanto elementos-chave indicados pelas normas internacionais para o estabelecimento de interoperabilidade

semântica, advém da inacessibilidade destes documentos, por fatores principalmente idiomáticos e financeiros.

A reflexão sobre os mapeamentos semânticos, proporcionada pela investigação realizada com as bibliotecas participantes, nos indica que os mapeamentos semânticos podem ser vistos como uma forma de validação da linguagem de indexação construída. Dado o estabelecimento de equivalências e demais relacionamentos com outras linguagens é possível conceber que esta linguagem seja enriquecida. Pesquisas futuras neste tópico poderiam abordar o mapeamento semântico como um tipo de garantia, tais como as que a literatura de organização do conhecimento reconhece atualmente.

Como recomendações às investigações futuras neste tema, sugere-se também a ampliação de bibliotecas participantes, aumentando a quantidade de projetos de linguagens de indexação analisados, que dessa maneira permita identificar experiências em que efetivamente estejam sendo utilizadas as diretrizes propostas na normalização internacional quanto a interoperabilidade semântica, deste modo, as recomendações propostas aqui poderiam ser também avaliadas e validadas.

## REFERÊNCIAS

- AMORIM, Igor Soares; CAFÉ, Lígia Maria Arruda. Os conceitos de comunidade discursiva, domínio e linguagem na análise de domínio Hjørlandiana. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (XVII ENANCIB)*, 17., 2016, Salvador. Anais [...]. Salvador: PPGCI, UFBA, 2016. Disponível em: <http://hdl.handle.net/123456789/3377>. Acesso em: 30 jul. 2020.
- ANDRADE, Juliatti de. **Interoperabilidade e mapeamentos entre sistemas de organização do conhecimento na busca e recuperação de informações em saúde: estudo de caso em ortopedia e traumatologia**. 2015. 327f. Tese (Doutorado em Ciência da Informação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2015. Disponível em: <https://www.teses.usp.br/teses/disponiveis/27/27151/tde-29062015-121813/en.php>. Acesso em: 30 jul. 2020.
- ANDRADE, Juliatti de; LARA, Marilda L. G. de. Interoperability and mapping between Knowledge Organization Systems. **Knowledge Organization**, Frankfurt, v. 43, n. 2, p. 107-112, 2016. Disponível em: <https://bit.ly/311ZV4u>. Acesso em: 30 jul. 2020.
- ANDRADE, Morgana Carneiro. **A interoperabilidade semântica na perspectiva da organização do conhecimento: uma proposta para o Repositório Institucional da Universidade Federal do Espírito Santo**. 2012. 212f. Dissertação (Mestrado Profissional em Gestão da Informação) – Programa de Pós-Graduação em Gestão da Informação, Centro de Educação, Comunicação e Artes, Universidade Estadual de Londrina, Londrina, 2012. Disponível em: <http://repositorio.ufes.br/handle/10/2021>. Acesso em: 30 jul. 2020.
- ANDRADE, Morgana Carneiro. **Interoperabilidade semântica entre repositórios digitais científicos: um perfil de aplicação de metadados baseado nas necessidades de informação em oncologia**. 2018. 431f. Tese (Doutorado) – Programa de Doutorado em Tecnologias e Sistemas de Informação, Universidade do Minho, Guimarães, Portugal, 2018. Disponível em: <https://repositorium.sdum.uminho.pt/handle/1822/58953>. Acesso em: 27 maio 2020.
- ANDRADE, Morgana Carneiro; CERVANTES, Brígida Maria Nogueira. A contribuição da organização do conhecimento para a interoperabilidade semântica: alternativas para repositórios institucionais. **Informação@ Profissões**, Londrina, v. 1, n. 1/2, p. 151-169, 2013. DOI: <http://dx.doi.org/10.5433/2317-4390.2012v1n1/2p151>. Disponível em: <http://dx.doi.org/10.5433/2317-4390.2012v1n1/2p151>. Acesso em: 30 jul. 2020.
- AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION. ANSI/NISO Z39.19-2005: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda, Ma: NISO Press, 2005. 184 p.
- ANSI/NISO Z39.50 - INFORMATION retrieval (Z39.50): application service definition and protocol specification. Bethesda, Ma: NISO, 1995. 155p
- ARAVE, G.; JACOB, E. K. Evaluating semantic interoperability across ontologies. *In: Knowledge organization for a sustainable world: challenges and perspectives for cultural,*

scientific, and technological sharing in a connected society, XIV, 2016. **Proceedings of the Fourteenth International ISKO Conference**, Würzburg: Ergon-Verlag, 2016, p. 308-316. Disponível em: <https://repositorium.sdum.uminho.pt/handle/1822/58953>. Acesso em: 30 jul. 2020.

AUSTIN, D. An Indexing manual for PRECIS. **International Classification**. v.1, n. 2, p. 91-94, 1974.

AUSTIN, Derek; DALE, Peter. **Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngües**. Brasília: Ibict/SENAI, 1993.

BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2011.

BARITE, Mario. Sistemas de organización del conocimiento: una tipología actualizada. **Informação & Informação**, Londrina, v. 16, n. 3, p. 122-139, 2011.

DOI: <http://dx.doi.org/10.5433/1981-8920.2011v16n2p122>. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2011v16n2p122>. Acesso em: 30 jul. 2020.

BOCCATO, Vera Regina C.; FUJITA, Mariângela Spotti Lopes. O uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: um estudo de avaliação sociocognitiva com protocolo verbal. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 15, n. 3, p. 23-51, 2006. DOI: <https://doi.org/10.1590/S1413-99362010000300003>. Disponível em: <https://doi.org/10.1590/S1413-99362010000300003>. Acesso em: 30 jul. 2020.

BOCCATO, Vera Regina C.; TORQUETTI, M. C. Interoperabilidade entre linguagens de indexação como recurso de construção de instrumento de representação temática de clippings de coordenadorias de comunicação social em ambientes universitários: uma proposta metodológica. **Informação & Informação**. v. 17, n. 3, p. 76-101, set./dez. 2012. DOI: <http://dx.doi.org/10.5433/1981-8920.2012v17n3p76>. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/10800>. Acesso em: 30 jul. 2020.

BRUYNE, P. de; HERMAN, Jacques; SCHOUTHEETE, M. de. **Dinâmica da pesquisa em ciências sociais: os polos da prática metodológica**. 5. ed. Rio de Janeiro: Francisco Alves, 1991.

BRITISH STANDARDS INSTITUTION; BS 8723-1:2005 Structured Vocabularies for Information Retrieval– Guide – Definitions, Symbols, and Abbreviations. London: British Standards Institution; 2005.

BRITISH STANDARDS INSTITUTION; BS 8723-2:2005 Structured Vocabularies for Information Retrieval– Guide – Thesauri. London: British Standards Institution; 2005.

CAMPOS, Maria Luiza de Almeida; GOMES, Hagar Espanha. Metodologia de elaboração de tesouro conceitual: a categorização como princípio norteador. **Perspectivas em Ciência da Informação**, v. 11, n. 3, p. 348-359, 2006. DOI: <https://doi.org/10.1590/S1413-99362006000300005>. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1413-99362006000300005&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362006000300005&lng=en&nrm=iso). Acesso em: 30 jul. 2020.

CAMPOS, Maria Luiza Almeida et al. Estudo comparativo de softwares de construção de tesouros. **Perspectivas em ciência da informação**, [S. l.], v. 11, n. 1, 2006. DOI: <https://doi.org/10.1590/S1413-99362006000100006>. Disponível em: [https://www.scielo.br/scielo.php?script=sci\\_arttext&%20pid=S1413-99362006000100006&lng=en&nrm=iso&tlng=pt](https://www.scielo.br/scielo.php?script=sci_arttext&%20pid=S1413-99362006000100006&lng=en&nrm=iso&tlng=pt). Acesso em: 30 jul. 2020.

CARNEIRO, M. V. Diretrizes para uma política de indexação. **Revista da escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 221-241, set. 1985.

CERVANTES, Brígida Maria Nogueira. **Contribuição para a terminologia do processo de inteligência competitiva: estudo teórico e metodológico**. 2004. 183f. Dissertação (Mestrado em Ciência da informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2004. Disponível em: <http://hdl.handle.net/11449/93723>. Acesso em: 10 ago. 2020.

CERVANTES, B. M. N.; FUJITA, Mariângela Spotti Lopes. Uma perspectiva diacrônica de parâmetros e modelos de construção de tesouros. *In.*: CERVANTES, B. M. N. (Org.). **Horizontes da organização da informação e do conhecimento**. Londrina: EDUEL. v. 1, p. 1-18.

COATES, E. J. **Subject catalogues: headings and structures**. London: Library Association, 1960.

CRUZ, Maria Carolina Andrade e; SANTOS, Luciana Beatriz Piovezan dos; FUJITA, Mariângela Spotti Lopes. Linguagens de indexação em bibliotecas universitárias brasileiras: diagnóstico preliminar das Regiões Sul e Sudeste. *In.*: ENCONTRO DA ASSOCIAÇÃO DE EDUCAÇÃO E PESQUISA EM CIÊNCIA DA INFORMAÇÃO DA IBERO-AMÉRICA E CARIBE, 10., 2016, Belo Horizonte. **Anais [...]**. Belo Horizonte: ECI/UFMG, 2016.

CURRÁS, Emília. **Ontologías, taxonomías y tesouros**. Gijón: Trea, 2005.

DAHLBERG, Ingetraut. Teoria do conceito. **Ciência da informação**, Brasília, v. 7, n. 2, 1978. Disponível em: <http://revista.ibict.br/ciinf/article/view/115>. Acesso em: 30 jul. 2020.

DAHLBERG, Ingetraut. Toward establishment of compatibility between indexing languages. **International Classification**, v. 8, n. 2, p. 86-91, 1981. Disponível em: [https://www.ergon-verlag.de/isko\\_ko/downloads/ic\\_8\\_1981\\_2\\_e.pdf](https://www.ergon-verlag.de/isko_ko/downloads/ic_8_1981_2_e.pdf). Acesso em: 30 jul. 2020.

DEXTRE CLARKE, Stella G.; ZENG, Marcia Lei. From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modeling. **Information Standards Quarterly**, [S. l.], v. 24, n. 1, 2012. Disponível em: <http://eprints.rclis.org/16818/>. Acesso em: 16 out. 2017.

DIRECTRICES para registros de autoridad y referencias. 2 ed. rev. por el Grupo de Trabajo de la IFLA para la Revisión de las GARE. 2001. Disponível em: <http://www.ifla.org/files/assets/cataloguing/garr/garr-es.pdf>. Acesso em: 30 jul. 2020.

DU, W. et al. Establishing interoperability among knowledge organization systems for research management: a social network approach. **Scientometrics**, [S. l.], v. 112, n. 3, p.

1489– 1506, 2017. Disponível em: <https://link.springer.com/article/10.1007/s11192-017-2457-0>. Acesso em: 15 Jan. 2018.

EL HADI, Widad Mustafa. Cultural Interoperability and Knowledge Organization Systems. *In: José Augusto Chaves Guimarães; Vera Dodebei. (Org.). **Organização do conhecimento e diversidade cultural***. Marília: ISKO-Brasil; FUNDEPE, v. 1, 2015. p. 575-606. *E-book*.

FOSKETT, A. C. Pesquisa na recuperação da informação. **A abordagem temática da informação**. São Paulo: Polígono, 1973.

FUJITA, Mariângela Spotti Lopes. Organização do conhecimento: algumas considerações para o tratamento temático da informação. *In: CARRARA, Kester. (Org.). **Educação, universidade e pesquisa***. Marília: Unesp-Marília-Publicações; São Paulo: FAPESP, 2001. p. 29-34.

FUJITA, Mariângela Spotti Lopes. Modelos de categorização para a construção de tesouros: metodologia de ensino. *In: BOCCATO, Vera Regina C.; GRACIOSO, L. de S. **Estudos de linguagem em Ciência da Informação***. Campinas: Alínea, 2011.

FUJITA, Mariângela Spotti Lopes. A política de indexação para representação e recuperação da informação. *In: GIL LEIVA, Isidoro; FUJITA, Mariângela Spotti Lopes. **Política de indexação***. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012. p. 17-28. *E-book*.

FUJITA, Mariângela Spotti Lopes., Projeto de pesquisa 2015.

FUJITA, Mariângela Spotti Lopes; PIOVEZAN, Luciana Beatriz; SANTOS, Natany Souza. A função do registro de autoridade de assunto na construção e uso de linguagens de indexação para catálogos online. *In: XIII CONGRESSO ISKO ESPANHA; III CONGRESSO ISKO ESPANHA-PORTUGAL, 2017, Coimbra. **Tendências atuais e perspectivas futuras em organização do conhecimento**: atas do III Congresso ISKO*. Coimbra: Universidade de Coimbra. Centro de Estudos Interdisciplinares do Século XX - CEIS20, 2017, v. 1, p. 577-586.

FUJITA, Mariângela Spotti Lopes; SANTOS, Luciana Beatriz Piovezan. A estrutura lógico-hierárquica de linguagens de indexação utilizadas por bibliotecas universitárias. *Scire*, Zaragoza, v. 22, n.2 (jul.-dic. 2016), 37-46, 2016. Disponível em: <https://ibersid.eu/ojs/index.php/scire/article/view/4336>. Acesso em: 02 out. 2018.

FUJITA, Mariângela Spotti Lopes; SANTOS, Luciana Beatriz Piovezan. Política de indexação em bibliotecas universitárias: estudo diagnóstico e analítico com pesquisa participante. *Transinformação*, Campinas, v. 28, p. 59-76, 2016. DOI: <http://dx.doi.org/10.1590/2318-08892016002800005>. Disponível em: <http://ref.scielo.org/dwzc8n>. Acesso em: 02 out. 2018.

FUJITA, Mariângela Spotti Lopes; SANTOS, Luciana Beatriz Piovezan dos; ALVES, Roberta Vesú. Linguagem de indexação e linguagem documentária são sistemas de organização do conhecimento? Uma análise bardiana da variação terminológica. *Scire*, Zaragoza, v. 24, n. 2, p. 23-33, 2018. Disponível em: <http://www.ibersid.eu/ojs/index.php/scire/article/view/4577>. Acesso em: 06 abr. 2019.

FUJITA, Mariângela Spotti Lopes et al. Avaliação das características do TemaTres e Multites para o controle de autoridades nas bibliotecas universitárias. **Scire**, Zaragoza, v. 23, n. 2, p. 63-73, 2017. Disponível em: <https://www.ibersid.eu/ojs/index.php/scire/article/view/4430>. Acesso em: 30 jul. 2020.

FUJITA, Mariângela Spotti Lopes et al. Construction and evaluation of hierarchical structures of indexing languages for online catalogs of libraries: an experience of the São Paulo State University (UNESP). **Knowledge Organization**, v. 45, n. 3, p. 220-231, 2018. DOI: <https://doi.org/10.5771/0943-7444-2018-3>. Disponível em: <https://www.nomos-elibrary.de/10.5771/0943-7444-2018-3-220/construction-and-evaluation-of-hierarchical-structures-of-indexing-languages-for-online-catalogs-of-libraries-an-experience-of-the-sao-paulo-state-university-unesp-volume-45-2018-issue-3>. Acesso em: 30 jul. 2020.

FUJITA, Mariângela Spotti Lopes et al. Linguagens de indexação em bibliotecas universitárias: estudo analítico. **Informação & Informação**, Londrina, v. 24, n. 1, p. 190-225, 2019. DOI: <http://dx.doi.org/10.5433/1981-8920.2019v24n1p190>. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/31771>. Acesso em: 06 abr. 2019.

GARCÍA MARCO, Francisco Javier. ISO 25964-2: interoperabilidad de los tesauros con otros sistemas de organización del conocimiento en contextos semánticos. **Anuario ThinkEPI**, [S. l.], n. 1, p. 305-309, 2017. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=6008488>. Acesso em: 7 nov. 2017.

MARCO, Francisco Javier Garcia. The evolution of thesauri and the history of knowledge organization: between the sword of mapping knowledge and the wall of keeping it simple. **Brazilian Journal of Information Science: Research Trends**, Marília, v. 10, n. 1, 2016. DOI: <https://doi.org/10.36311/1981-1640.2016.v10n1.01.p1>. Disponível em: <http://revistas.marilia.unesp.br/index.php/bjis/article/view/5786>. Acesso em: 30 jul. 2020.

GIL LEIVA, Isidoro. Consistencia en la asignación de materias en bibliotecas públicas del Estado. **Boletín de la Asociación Andaluza de Bibliotecarios**, [S. l.], n.63, p.69-96, 2001. Disponível em: <http://eprints.rclis.org/5963/>. Acesso em: 30 jul. 2020.

GIL LEIVA, Isidoro. **Manual de indización**. Gijón: Trea, 2008.

GIL URDICIAIN, B. **Manual de lenguajes documentales**. 2. ed. rev. e aum. Madrid: TREA, 2004.

GOLUB, Koraljka. Knowledge Organisation Systems. **Technical Foundations**, UK. 6. 2011. Disponível em: <https://www.diva-portal.org/smash/get/diva2:793321/FULLTEXT01.pdf>. Acesso em: 30 jul. 2020.

GOMES, Hagar Espanha. **Manual de elaboração de tesauros monolíngues**. Brasília: Programa Nacional de Bibliotecas das Instituições de Ensino Superior, 1990.

GÓMEZ DUEÑAS, Laureano Felipe. **Modelos de interoperabilidad en bibliotecas digitales y repositorios documentales: caso Biblioteca Digital Colombiana BDCOL**. Disponível em: <http://eprints.rclis.org/14878/>. Acesso em: 30 jul. 2020.

HOSHY, L. M. E. Relationships in Library of Congress Subject Headings. In.: BEAN, C. A. (ed.) **Relationships in the organization of knowledge**. Dordrecht: Kluwer Academic Publishers, 2001. 135-152.

HODGE, G. **Systems of knowledge organization for digital libraries: beyond traditional authorities' files**. Washington, DC: CLIR, 2000. Disponível em: <http://www.clir.org/pubs/reports/pub91/contents.htm>. Acesso em: 30 Jul. 2020.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 2788-1986. Documentation – Guidelines for the Establishment and Development of Monolingual Thesauri. 2nd ed. Geneva: International Organization for Standardization; 1986.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 25964-1: information and documentation - Thesauri and interoperability with other vocabularies - part 1: Thesauri for information retrieval. Genebra, 2011.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 25964-2: information and documentation - Thesauri and interoperability with other vocabularies - part 2: Interoperability with other vocabularies. Genebra, 2013.

LANCASTER, Frederick W. **El control del vocabulario en la recuperación de información**. Universitat de València, 1995.

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

LEIVA MEDEROS, Amed et. al. Working framework of semantic interoperability for CRIS with heterogeneous data sources. **Journal of Documentation**, [S. l.], v. 73, n. 3, p. 481-499, 2017. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/JD-07-2016-0091/full/html>. Acesso em: 30 jan. 2018.

LLANES PADRÓN, Dunia; FUJITA, Mariângela Spotti Lopes; BASTOS, Flávia Maria. Os registros de autoridade em sistemas de informação: uma perspectiva biblioteconômica e arquivística. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (XV ENANCIB), 15., 2014, Belo Horizonte. **Anais** [...]. Belo Horizonte: ECI/UFMG, 2014. Disponível em: <http://hdl.handle.net/123456789/2606>. Acesso em: 30 jul. 2020.

LLANES-PADRÓN, Dunia; PASTOR-SÁNCHEZ, Juan-Antonio. Records in contexts: the road of archives to semantic interoperability. **Program: electronic library and information systems**, [S. l.], v. 51, n. 4, p. 387-405, 2017. DOI: <https://doi.org/10.1108/PROG-03-2017-0021>. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/PROG-03-2017-0021/full/html>. Acesso em: 30 jul. 2020.

LUCARELLI, Anna; VITI, Elisabetta. Florence–Washington Round Trip: Ways and Intersections between Semantic Indexing Tools in Different Languages. **Cataloging & Classification Quarterly**, [S. l.], v. 53, n. 3-4, p. 414-429, 2015. Disponível em: <https://doi.org/10.1080/01639374.2014.1004008>. Acesso em: 22 fev. 2019.



- MAI, Jens-Erik. The future of general classification. **Cataloging & classification quarterly**, [S. l.], v. 37, n. 1-2, p. 3-12, 2003. DOI: [https://doi.org/10.1300/J104v37n01\\_02](https://doi.org/10.1300/J104v37n01_02). Disponível em: [https://www.tandfonline.com/doi/abs/10.1300/J104v37n01\\_02](https://www.tandfonline.com/doi/abs/10.1300/J104v37n01_02). Acesso em: 30 jul. 2020.
- MARTÍNEZ-GONZÁLEZ, M. Mercedes; ALVITE-DÍEZ, María-Luisa. On the evaluation of thesaurus tools compatible with the Semantic Web. **Journal of Information Science**, v. 40, n. 6, p. 711-722, 2014. DOI: <https://doi.org/10.1177/0165551514545603>. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0165551514545603>. Acesso em: 30 jul. 2020.
- MARTÍNEZ TAMAYO, Ana María; VALDEZ, Julia Catalina. **Indización y clasificación en bibliotecas**. Buenos Aires: Alfagrama, 2008.
- MARTINS, Maria Catarina Barradas. **Indexação e controlo da terminologia em bibliotecas do ensino superior politécnico em Portugal: o sistema no instituto politécnico de Portalegre**. Salamanca: Ediciones Universidad de Salamanca, 2013.
- MAYR, P. et al. Recent applications of Knowledge Organization Systems: introduction to a special issue. **International Journal on Digital Libraries**, [S. l.], v. 17, n. 1, p. 1-4, 2016. DOI: <https://doi.org/10.1007/s00799-015-0167-x>. Disponível em: <https://link.springer.com/content/pdf/10.1007/s00799-015-0167-x.pdf>. Acesso em: 30 Jul. 2020.
- MAZZOCCHI, Fulvio. Relations in KOS: is it possible to couple a common nature with different roles? **Journal of Documentation**, [S. l.], v. 73, n. 2, p. 368-383, 2017. Disponível em: <https://doi.org/10.1108/JD-05-2016-0063>. Acesso em: 25 nov. 2017.
- MELO, Maria Antônia Fonseca; BRÄSCHER, Marisa. Requisitos Funcionais para Dados de Autoridade Assunto (FRSAD): entidades, atributos e relacionamentos. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, [S. l.], v. 12, n. 2, p. 102-119, 2014.
- MENDES, Paula Raphisa; REIS, RM dos; MACULAN, BCM dos S. Tesouros no acesso à informação: uma retrospectiva. **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis, v. 20, n. 1, p. 49-66, 2015.
- MOREIRA, Manoel Palhares; MOURA, Maria Aparecida. Construindo tesouros a partir de tesouros existentes: a experiência do TCI-Tesouro em Ciência da Informação. **DataGramZero: Revista de Ciência da Informação**, Brasília, v. 7, n. 4, ago. 2006. Disponível em: [https://brapci.inf.br/\\_repositorio/2010/01/pdf\\_6c43aff315\\_0007598.pdf](https://brapci.inf.br/_repositorio/2010/01/pdf_6c43aff315_0007598.pdf). Acesso em: 30 jul. 2020.
- MOREIRA, Walter. Tesouros e ontologias como modelos de sistemas de organização do conhecimento. **Brazilian Journal of Information Science: research trends**, Marília, v. 13, n. 1, p. 15-20, 2019. DOI: <https://doi.org/10.36311/1981-1640.2019.v13n1.03.p15>. Disponível em: <https://www2.marilia.unesp.br/index.php/bjis/article/view/8277>. Acesso em: 30 jul. 2020.
- MOREIRA, Walter; LARA, Marilda L. G. de. Ontologias, categorias e interoperabilidade semântica. **DataGramZero: revista de informação**, Brasília, v. 13, n. 4, ago 2012. Disponível em: [http://profwalter.com.br/meusArtigos/%5B2012,DataGramZero%5DMoreira,W\\_OntologiasCategoriasEInteroperabilidadeSemantica.pdf](http://profwalter.com.br/meusArtigos/%5B2012,DataGramZero%5DMoreira,W_OntologiasCategoriasEInteroperabilidadeSemantica.pdf). Acesso em: 30 jul. 2020.

NEWMAN, S. M. **Information systems compatibility**. Washington: Spartan Books; London: Macmillan, 1965.

NEVILLE, H. H. Thesaurus reconciliation. **Aslib Proceedings**, [S. l.], v. 11, n. 24, nov. 1972, p. 620-626, 1972. DOI: <https://doi.org/10.1108/eb050379>. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/eb050379/full/html>. Acesso em: 30 jul. 2020.

OLIVEIRA, Henry Poncio Cruz de. **Arquitetura da informação pervasiva: contribuições conceituais**. 2013. 203 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013. Disponível em: <http://hdl.handle.net/11449/110387>. Acesso em: 30 jul. 2020.

OLSON, Hope A.; BOLL, John J. **Subject analysis in online catalogs**. 2. ed. Englewood, CO: Libraries Unlimited, 2001.

PASTOR SANCHEZ, J. A.; MARTÍNEZ MÉNDEZ, F. J.; RODRÍGUEZ MUÑOZ, J. V. Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de *linked open data*. **El Profesional de la Información**, [S. l.], v. 21, n. 3, p. 245-253, 2012. Disponível em: <http://hdl.handle.net/10760/17066>. Acesso em: 30 jul. 2020.

PIOVEZAN, Luciana Beatriz. **Avaliação da indexação em catálogos de bibliotecas universitárias por meio da recuperação da informação**. 2015. 103f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2015.

REQUISITOS Funcionales para Datos de Autoridad de Materia (FRSAD). Grupo de Trabajo de IFLA sobre los Requisitos Funcionales para Registros de Autoridad de Materia (FRSAR). Disponível em: <http://www.ifla.org/files/assets/cataloguing/frsad/frsad-final-report-es.pdf.pdf>

RUBI, Milena Polsinelli. Política de indexação. In: GIL LEIVA, Isidoro.; FUJITA, Mariângela Spotti Lopes. **Política de indexação**. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012. p. 107-120. *E-book*.

SAN SEGUNDO, R. **Sistemas de organización del conocimiento: la organización del conocimiento en las bibliotecas españolas**. Madrid: Universidad Carlos III de Madrid, Boletín Oficial del Estado, 1996.

SANTOS, Luciana Beatriz Piovezan dos. **Política de indexação em bibliotecas universitárias: estudo diagnóstico na região de Marília-SP**. 2011. 58f. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2011.

SAYÃO, Luis Fernando. Padrões para bibliotecas digitais abertas e interoperáveis. **Encontros Bibli**, Florianópolis, n. esp., 1. sem. 2007. Disponível em: <https://doi.org/10.5007/1518-2924.2007v12nesp1p18>. Acesso em: 02 out. 2018.

SILVA, Armando Malheiro da. O método quadripolar e a pesquisa em ciência da informação. **Prisma.com**, n. 26, p. 27-44, 2014.

SILVA, A. M.; RIBEIRO, F. **Das “ciências” documentais à ciência da informação**: ensaio epistemológico para um novo modelo curricular. Porto: Edições Afrontamento, 2002.

SOERGEL, Dagobert. Indexing and retrieval performance: the logical evidence. **Journal of the American Society for Information Science**, Silver Spring, MD, v. 45, n. 8, p. 589-599, 1994.

VAN SLYPE, Georges. **Lenguajes de indización**: concepción, construcción y utilización en los sistemas documentales. Madrid: Fundación Germán Sánchez Rui Perez, 1991.

WELLISCH, H. H. **Indexing from A to Z**. 2. ed. New York: Wilson, 1995.

## APÊNDICE A – TRANSCRIÇÃO DE ENTREVISTA I

**Local:** Biblioteca FCL-Ar Unesp

**Duração:** 13:38 min

*Pesquisadora:* Você leu o termo de consentimento? Está tudo ok para você?

*Sujeito:* Sim, tá tudo ok.

*Pesquisadora:* Tá certo, então. Eu vou fazendo as perguntas e vou digitando algumas anotações com relação às suas respostas, tá? Também pra ficar mais fácil.

*Sujeito:* hum hum

*Pesquisadora:* A primeira pergunta é sobre você mesmo, para você se identificar e a biblioteca que você trabalha. [1. Identificação da biblioteca e respondente:]

*Sujeito:* Meu nome é Aline Aparecida Matias e eu trabalho na Faculdade de Ciências e Letras de Araraquara.

*Pesquisadora:* Certo. E qual é a função que você desempenha hoje na biblioteca? [2. Respondente desempenha qual função na biblioteca?]

*Sujeito:* No momento eu sou supervisora da seção, da STATI, da Seção de Tratamento e Aquisição da Informação.

*Pesquisadora:* Sim. E você também faz catalogação, nessa função?

*Sujeito:* Sim, faço.

*Pesquisadora:* Ah, sim. E junto na STATI, quantas pessoas têm fazendo catalogação hoje?

*Sujeito:* Eu tenho mais três bibliotecárias, duas que fazem catalogação de livros, de teses, DVDs e uma que trabalha só com periódicos.

*Pesquisadora:* Mais três, então, né? Você e mais três, são quatro pessoas fazendo catalogação hoje em dia na biblioteca.

*Sujeito:* É, hum hum.

*Pesquisadora:* Ok. A terceira pergunta é se a biblioteca desenvolve ou desenvolveu projeto construção ou manutenção/atualização de linguagem de indexação com interoperabilidade semântica?

*Sujeito:* Na verdade a biblioteca separadamente não, né, a gente como faz parte da rede de bibliotecas Unesp então esse projeto é desenvolvido de forma coletiva.

*Pesquisadora:* Ok. E qual software está sendo utilizado para desenvolver a linguagem?

*Sujeito:* Olha, a gente o que a gente fica sabendo, que é o que a gente trabalha, a gente trabalha como a linguagem integrada dentro do Aleph, o que é o software que a gente usa para fazer catalogação.

*Pesquisadora:* Humhum. Ok.

*Sujeito:* Porque às vezes, quando é necessário fazer, por exemplo, catalogação de autor, de assunto, a gente já faz diretamente no sistema.

*Pesquisadora:* Sim, sim, ok. E sobre o formato de metadados foi adotado para desenvolver a linguagem?

*Sujeito:* Ah, a gente usa o mesmo formato de metadados que a gente usa para catalogação dos livros, mas claro usando os campos específicos do MARC no que diz respeito a catalogação de assunto ou de autor.

*Pesquisadora:* É o MARC Autoridades?

*Sujeito:* É, MARC Autoridades.

*Pesquisadora:* A próxima pergunta é se você nesse trabalho se as linguagens que estão sendo usadas para construir a linguagem, se elas têm coordenação diferentes, algumas são pré-coordenadas, outras pós-coordenadas? Como, por exemplo, a CDD e CDU são pré-coordenadas, a LCSH, pós-coordenada? [6. Está trabalhando com linguagens que possuem modos de coordenação diferentes? Ex. PrêC- CDD/CDU e PósC – LCSH, tesouros...].

*Sujeito:* A questão é que, na hora de fazer os cabeçalhos, na hora de pensar nessa construção, a gente acaba usando essa que são pós-coordenadas, que é a LC que é BN, que você tem essa liberdade para construir os cabeçalhos. No caso aqui, na classificação a gente usa a CDD, então são... eles se complementam, mas são coisas diferentes.

*Pesquisadora:* Então, acabam usando as duas, mas para funções diferentes?

*Sujeito:* Exatamente. A CDD a gente usa para identificação do material na estante, para composição do número na etiqueta, para facilitar para o aluno encontrar, para a gente conseguir ordenar o material e, a linguagem própria, a linguagem documentária a gente usa para conseguir expressar o conteúdo do documento da forma mais eficaz possível.

*Pesquisadora:* Sim, sim. E para a fusão das linguagens, linguagens pré-coordenadas não têm sido utilizadas, só linguagens pós-coordenadas?

*Sujeito:* Não. Sim, sim.

*Pesquisadora:* Ok. E você, no seu cotidiano, percebeu alguma dificuldade nesta questão da coordenação?

*Sujeito:* Não.

*Pesquisadora:* ok. E você percebeu se têm sido utilizadas linguagens com níveis de especificidade diferentes para fazer a fusão das linguagens, linguagens mais gerais *versus* linguagens mais específicas, como às vezes as linguagens da área de assunto, tesouros de áreas específicas? [8. Está trabalhando com linguagens que possuem níveis de especificidade diferentes? Ex. Geral: TBN, LCSH e Específicas: MeSH, DeCS, Tesouros de áreas, Agrovoc etc.]

*Sujeito:* Ah, é o seguinte aqui a nossa área específica é a área de Ciências Humanas, então a gente tem Letras, Pedagogia, Ciências Econômicas, Administração e Ciências Sociais. Nesses casos, é mais fácil a gente encontrar mesmo na própria linguagem. Quando a gente precisa fazer buscas externas, sejam na LC ou na BN, a gente consegue encontrar mais os termos que a gente precisa, pois são termos assim mais consolidados, e às vezes, são bem específicos. Aí eu não sei te dizer como seria a experiência no caso de quem trabalha numa biblioteca que acontecem atualizações mais... assim, mais recorrentes, como uma biblioteca da área de saúde, que é o que eu vejo às vezes quando a gente conversa com o pessoal, que eles dizem que é complicado, que tem termos que eles precisam e que não tá disponível em nenhum lugar. [Não.]

*Pesquisadora:* Então, aí no seu caso como é a área mais voltada para as Humanas, em geral, os termos você encontra nas linguagens mais gerais, como as grandes linguagens?

*Sujeito:* Isso exatamente, são termos mais consolidados. São raras as exceções, eu acho que até hoje a gente precisou uma única vez solicitar a criação de um termo que não tinha nem na LC nem na BN e aí a gente passou por todo aquele processo de preencher um formulário, de pedir ajuda ao docente específico da área, de levantar fontes, mas foi uma única vez, nas outras vezes que a gente precisou de assuntos eles estavam contemplados ou na BN ou na própria LC, se já não estavam na própria linguagem da UNESP.

*Pesquisadora:* Perfeito. Então esse ponto não é uma dificuldade para vocês?

*Sujeito:* Não, não, de maneira nenhuma.

*Pesquisadora:* Ok. Então a última pergunta já.

*Sujeito:* Ah, já tá acabando!

*Pesquisadora:* É são poucas perguntas. Se nesse processo de construção, em geral, principalmente com relação aí dessas, o uso das três linguagens diferentes se vocês perceberam alguma dificuldade nesse processo? Em geral, qualquer dificuldade. [Qual outra dificuldade foi encontrada durante o processo de desenvolvimento/atualização da linguagem com relação à interoperabilidade semântica?]

*Sujeito:* Às vezes, por exemplo, uma das coisas que a gente achou bem assim diferente, foi quando foi criar os cabeçalhos para personagens fictícios, quando você usa o nome da pessoa seguido do personagem fictício, por exemplo, Robson Crusóé aí entre parênteses personagem fictício. A BN tratava ele num campo e a LC em outro completamente diferente, um tratava como assunto tópico e a outra tratava como nome pessoal. E aí era uma diferença que no começo a gente ficou “Como que a gente faz?” e aí quem acabou decidindo foi o grupo né, que trata de indexação, e aí fizeram reuniões... e aí decidiram como é, para ficar padronizado né, por que às vezes a pessoa, uma de nós da rede importava da BN e ficava num determinado campo e quem importava da LC ficava em outro e era a mesma coisa.

[Criar cabeçalhos para personagens fictícios, tem dificuldades com uso do formato de metadados, em uma linguagem utilizadas para fusão interpreta de uma forma e outra linguagem o campo é outro, dificulta o mapeamento.]

*Pesquisadora:* Entendi. E aí você acha o que é uma dificuldade mais da forma como essas duas linguagens colocam as informações nos metadados?

*Sujeito:* Exatamente. O entendimento que tem de se aquela pessoa, o fictício, ele é só um assunto tópico ou se, apesar de ser fictício, ele é uma pessoa sim e deveria ser tratado como um campo 100.

*Pesquisadora:* Entendi, entendi. Tem mais alguma coisa que você acha relevante colocar sobre toda essa questão da linguagem, de uso de linguagens diferentes para a construção da linguagem da Unesp?

*Sujeito:* Eu acho que foi, primeiro foi um passo muito importante a construção dessa linguagem, porque quando eu entrei aqui na Unesp, eu tô aqui desde janeiro de 2008, a gente fazia uso do cabeçalho autorizado do Bibliodata, só que às vezes aí o que acontecia, como também era uma coisa cooperativa, só que assim muito mais espalhada pelo país, então era complicado de construir e tinha a equipe no Rio, então o que que a gente fazia às vezes quando a gente procurava alguma coisa e era muito, por exemplo, na área de autor era muito, muito autor que a gente não encontrava, então o que a gente fazia, a gente fazia arquivos no Bloco de Notas e enviava uma vez por mês para a bibliotecária responsável lá do Bibliodata para construir. A

partir do momento em que a gente conseguiu construir o nosso próprio ... a nossa própria linguagem, Isso se tornou mais flexível, se tornou mais rápido, se tornou mais rápido tirar a dúvida, você mesmo já vai já e faz e fica cada vez mais evidente a característica principal, que é a cooperação na construção dessa linguagem. E é claro como a gente usa de linguagens diferentes, inclusive, além do que eu, aqui a gente só usa a BN e a LC, mas eu vejo nas instruções que vêm, por exemplo, o pessoal de saúde usa outras linguagens específicas, então é mais só a gente aparar essas pequenas arestas, mas acaba se tornando uma linguagem mais completa possível, pelo menos no meu ponto de vista.

*Pesquisadora:* Tá certo. Ah, isso me fez pensar em mais uma outra coisa, se você me permitir mais uma pergunta.

*Sujeito:* Claro. Fica à vontade.

*Pesquisadora:* Sobre as instruções que vocês na biblioteca recebem do grupo de linguagem, não só o manual, formalmente, mas também, em e-mails particulares e de instruções, se essas informações pra vocês têm sido úteis, se vocês voltam pra consultar?

*Sujeito:* Sim, sim, sim, por que a gente faz assim, a gente fez todos os treinamentos, e eu mais ainda que já estou aqui há esses 12 anos, eu acompanho assim desde quando começaram as primeiras reuniões. Então além de ter todo o material disponível lá e os treinamentos que a gente fez e as atualizações, sempre que a gente entra em contato com alguma dúvida, alguma necessidade de correção se for uma coisa simples, por exemplo, Ah eu vi na BN que o autor agora ele tem uma data de morte, aí eu faço essa solicitação, as meninas além de corrigir sempre manda uma devolutiva dizendo que corrigiram. Agora se é algum caso como foi esse do personagem, além de estabelecer como isso vai ser usado, a gente sempre recebe um feedback dizendo o porquê, ah a decisão foi essa então a gente faz assim e aí eu, por exemplo, quando eu recebo estagiárias todas essas informações que a gente têm, desde as que estão formalmente lá no site quanto essas que vão sendo trocadas por e-mail, que vão sendo respostas das nossas dúvidas, a gente vai passando, para deixar todo mundo com o mesmo nível de conhecimento do que tem que fazer no sistema.

*Pesquisadora:* Perfeito, Aline. Era isso o que eu tinha para te perguntar hoje, agradeço muito sua participação. E se depois, você tiver interesse em receber o material da pesquisa, para ver né, pode ser interessante para você, me fala, me manda um email que eu te dou essa devolutiva.

*Sujeito:* Ah, legal.



## APÊNDICE B - TRANSCRIÇÃO DE ENTREVISTA II

**Local:** – Instituto de Tecnologia da Aeronáutica

**Duração:** 1:07:35 min

*Pesquisadora:* Você leu o termo de consentimento? Está tudo ok para você?

*Sujeito 2:* Equipe então, termo de consentimento livre esclarecido eu faço ele numa folha separada, Ou eu imprimo com todas aquelas informações que antecederam o termo de Compromisso? Porque eu já assinei alguns e às vezes é com a folha, às vezes é sem a folha, rubrica a folha da frente. E eu confesso que não assinei e não encaminhei porque fiquei com dúvida.

*Pesquisadora:* Ah, tudo bem. Você pode preencher naquele arquivo mesmo que eu te enviei, e me enviar digitalizado.

*Sujeito 2:* Perfeito. Terminando a nossa entrevista eu faço e te encaminho.

*Pesquisadora:* Obrigada. Então eu vou começar as perguntas e eu vou anotando algumas coisas, para me ajudar também com a transcrição. Tudo bem?

*Sujeito 2:* Tudo bem, sem problemas.

*Pesquisadora:* Então a primeira pergunta, eu preciso da sua identificação e da biblioteca.

*Sujeito 2:* Identificação, consta do que? Nome, cargo, formação...

*Pesquisadora:* Isso, exatamente.

*Sujeito 2:* Eu sou Vera Porto Romeu Junqueira, formada pelas Faculdades integradas Teresa D'Ávila de Lorena, que agora é UNIFATEA. Sou registrada no CRB da 8ª região de São Paulo. Trabalho na biblioteca do ITA especificamente como bibliotecária Desde 1982 , trabalhei no processamento técnico, trabalhei na referência, Trabalhei na seção técnico administrativo, Quem gerencia a parte administrativa da biblioteca, desenvolvimento de coleções, Contato com fornecedores, E aí você entenda tudo, até a construção da biblioteca, Reformas, projetos. É atualmente estou chefe da biblioteca, desde 2008. Então, não é a minha primeira experiência como bibliotecária, mas é a de mais longa duração. Então, é biblioteca do ITA, conhecida como Divisão de Informação e Documentação – Biblioteca do ITA, hoje ligada à reitoria do ITA.

*Pesquisadora:* E qual função desempenha na biblioteca?

*Sujeito 2:* Ah, eu fui bibliotecária de referência, de processamento técnico, e uma questão importante, eu entrei na Biblioteca do ITA, bom não só eu, mas todo o corpo de bibliotecários

que entraram, foi para a automação dessa biblioteca, até outubro de 1982 ela fazia processos manuais e a partir de 17 de outubro de 1982 foi definido que tudo seria eletrônico. Então, nós usávamos um sistema antigo, preenchíamos formulários, hoje em dia todo mundo vê dá risada, mas digamos que na década de 80 isso era padrão, você tinha digitadores, hoje em dia fazemos tudo, o que é muito melhor e se comprovou com o menor número de inconsistências de recuperação de dados. E especificidade, os bibliotecários que entram nessa biblioteca basicamente fazem de tudo, então eles têm a especificidade, mas aprendem principalmente na referência a conhecer o acervo, depois aí a gente se desloca para as atividades que tem mais afinidade.

*Pesquisadora:* E sobre os processamentos técnico, você tem uma equipe?

*Sujeito 2:* Atualmente é uma equipe fechada. Hoje, em 2020 nós temos 12 funcionários, destes 12, 8 são bibliotecários, então, se você analisar ficou uma equipe fechada, mas o que acontece, qualquer um de nós, se houver demanda no processamento técnico vai para o processamento técnico. Então, o período de férias, qual é o período do ano mais crítico para a biblioteca do ITA para o processamento técnico, ele começa primeiro de outubro e termina a 31 de janeiro do ano subsequente, que é o período de entrega da produção técnico-científica. Basicamente, toda a produção técnico-científica do ITA é entregue aqui, mas literalmente o que é que pesa, os trabalhos de conclusão de curso, os trabalhos de graduação, as teses e as dissertações. Não é feito, diferentemente das outras instituições, não é feito o auto-arquivamento. Nós temos uma norma de produção técnico científica, que tem todo o processo e o último processo dela, depois da tese ou do trabalho defendido, é o depósito na biblioteca. Então ele faz um formulário de depósito, que é encaminhado para a equipe de processamento técnico, que faz uma pré-análise, a criação do objeto digital também a biblioteca presta esse suporte, se o texto original for em DOC, se ele usar Latex que já requer outro conhecimento, aí não é feito aqui, é a pós graduação que tem uma equipe que cuida disso para auxiliar. Então ele vem, a gente transforma o documento, faz os hiperlinks do documento, cria a chave porque precisa de segurança e aí define se o trabalho tem resultado passível de proteção, se a defesa foi em sigilo, isso consta no documento e ele é arquivado num banco paralelo. Então todo o conjunto, os trabalhos, os TGs como a gente chama, os trabalhos do curso de graduação, eles são digitais desde o ano de 2003, então eles já nasceram digital, de 2003 para frente eu já não tenho mais o papel também. As teses e dissertações, infelizmente, a gente ainda tem que cumprir a legislação da Capes, então a gente recebe um exemplar em papel e recebe o objeto digital, mas do próprio autor. Então tem todo um processo administrativo que o autor faz com a pós-graduação ou a graduação e a

biblioteca recebe o produto, mas nesse meio tempo ela tem o relacionamento com eles nos cursos de capacitação, ensinando como é que a biblioteca funciona, na capacitação de utilização das normas técnicas, na capacitação do uso das ferramentas, a qualquer momento grupos de alunos de pós-graduação que precisem de uma capacitação específica, eles podem solicitar e é dado essa capacitação a grupos de 2 ou 3 alunos. Então quando o trabalho chega a gente haveria de imaginar que já estava pronto, muitas vezes não, porque existem muitos ainda dos alunos que defendem a tese e aí se lembram que tem que olhar a formatação, as referências bibliográficas, então existem ainda alguns percalços, mas eles se restringem a menos de 10% do total de usuários. O ITA é uma instituição pequena, embora a gente tenha um reitor, se eu somar os cinco anos de graduação eu tenho 570 alunos e mais 1800 de pós-graduação. O que nós temos muito agora e que tem uma demanda muito grande são os cursos *lato sensu in company*, então nós temos cursos *lato sensu* no Itau, esse também a gente já definiu que o entregável é um artigo de periódicos baseado na NBR, ela dá o formato, ele apresenta este artigo de periódico, não necessariamente ele submete para um periódico, mas ele apresenta para uma banca, então o que é que me diz se esse trabalho tem credibilidade, se a parte técnica é confiável para eu depositar na biblioteca digital é essa banca que avaliou. E esse formulário de encaminhamento se chama formulário de registro, então ele já vem com essas informações.

*Pesquisadora:* Uma coisa que eu gostaria de saber é se você tem, e se você pode disponibilizar algum manual, alguma diretriz, relacionada aos processamentos técnico, essa questão da linguagem?

*Sujeito 2:* Olha nós temos algumas normas, algumas delas são de caráter ostensivo, não tem problema. O processamento técnico tem um manual insano, porque começou a ser construído em 82. Aí o que acontece em 82 você não tinha nem o ACCR2 traduzido para o português, então nós usávamos o AACR em inglês, nós usávamos as tabelas CDU, as extensões que também não estavam em português. Então a gente começou a construir vários manuais, então o manual de trabalho ele tem todas as, digamos assim, todos os ajustes que nós fizemos, porque de 2005 para cá nós temos um software proprietário, A gente adquiriu um software proprietário, mas de 82 a 2005 eram softwares desenvolvidos aqui. Então a gente usou o software que o INPE usava para, na realidade eu costumo brincar que ele levou alguns anos para ficar inteligente, a gente usava um grande computador, como máquina de escrever não é. Aí você dizia que o campo tal é esse o título, isso falando de uma maneira bem simplista, lógico que eu sei que tem coisas mais importantes do que isso por trás e obviamente foram esses dados que permitiram no futuro, hoje, as recuperações que eu tenho. Então eu tenho manual dessa época,

com as tradições que nós fizemos, os ajustes, nós usávamos, sempre usamos, os cabeçalhos de assunto da biblioteca do Congresso de Washington, Tesouros da NASA, Thesaurus do *Engineering joint Council*, usamos também o do IBICT, que ele tem um tesouro em português que foi feito em parceria seria com Portugal, vários tesouros para compor. A gente chegou a ter um Thesaurus Aeroespacial dos termos específicos da nossa área, a gente foi construindo a questão dos metadados, dos campos, a gente foi construindo com base, mas em alguma coisa muito sólida que já tinha. Então, nós tínhamos toda uma estrutura que precisava ser convertida agora para alguma coisa que a máquina entendesse, então nós fazemos ainda os processos à moda antiga, a gente cataloga, classifica e indexa e hoje isso tem uma única linguagem. E tem assim, algumas prerrogativas, a antiga chefe da seção, que entrou aqui comigo, ela tinha se aposentado em 2012 e aí eu consegui um projeto para trazê-la de volta por um ano, para que essas peculiaridades de sistema, como eu estava te dizendo, a gente pegou o sistema do INPE e teve que ir adaptando. Os sistemas hoje eles têm campo para tudo, tem MARC, tem tudo. Antigamente não era assim. O que a gente tinha era uma norma ISO que diz dado em rede de acervo, de informação bibliográfica, tem que ser passado dessa maneira, que é o que permite que os bancos se juntem e a gente consiga pesquisar. Então todo esse histórico, esse manual que eu estou te falando que tem isso, eu posso até conversar com ele e ver o que é o que é que a gente pode passar para você, porque tem muita coisa que é decisão.

*Pesquisadora:* E qual é o software que vocês utilizam?

*Sujeito 2:* É o Sophia, mas o nosso software é customizado, então tem muita gente que vem aqui vê o nosso Sophia e fala o meu não tem isso, o meu não tem aquilo e eu falo é porque é o nosso. E o Sophia foi um software eleito mediante aquela tabela de avaliação de softwares da Lígia Café, que era o instrumento que nós tínhamos a época. Então nós analisamos cinco pacotes de softwares de bibliotecas e o que mais atingiu os nossos objetivos foi o Sophia, que assim mesmo teve que ser customizado. Então algumas práticas a gente aprendeu, então a gente criou quando ele foi liberado uma máquina de homologação, então só libera para a produção a partir do momento que a gente homologa, então a gente fez um documento se tudo o que a gente pediu para fazer foi feito e tinham muitos ajustes que tinham que ser feitos. Porque nós já nascemos em rede em 82 e aí os sistemas eram diferentes não era como hoje, o código de barras é uma coisa que é mais recente, quando a gente começou a automação não se falava em código de barras, nos nossos primeiros registros. Então a gente é de um tempo que a impressora imprimia a ficha para você colocar no catálogo, então a gente soltava a ficha, alfabetava a ficha, nós tínhamos 16 catálogos, hoje a gente tem um só que ficou aqui por questões sentimentais. É

como nosso cardex está aqui por questões sentimentais, então nós vamos mandar restaurar para que tenha o quê a história desse processo. Então a avaliação do Sophia que a gente fez usando a ferramenta da Lígia Café eu acho que isso é importante. É um trabalho que foi um divisor de águas, para você ter uma ideia, o comando da aeronáutica tem o Centro de Documentação da Aeronáutica, CENDOC, que fica no Rio de Janeiro e que é responsável pela gestão de todas as bibliotecas e arquivos do Comando da Aeronáutica. Então você lembra do Oiapoque ao Chuí, eles é que cuidam disso. A biblioteca do ITA, o DCTA como um todo ele é muito peculiar, então é a maior biblioteca, é o maior acervo, é o maior tudo e a gente saiu na frente e nós conseguimos convencê-los com esse documento de usar um software único para todos. Então o que se precisa, aí me dizem, você prefere esse ou aquele software, eu falo não, nós não preferimos, não a gente insistiu naquele que respondeu as nossas expectativas técnicas. Então se tiver um outro no futuro, não tem problema nenhum para migrar desse para o outro, mas eles entenderam que a gente tem que ter um software único, então o CENDOC está comprando para as outras unidades. Então você tem a UNIFA, no Campo dos Afonsos, que é a Universidade da Força Aérea, usa o Sophia, a escola como a AFA, a Academia da Força Aérea, que forma os militares, também usa o Sophia, a EPCAR que forma os Cadetes, também usam o Sophia, aí você tem o DCEA, que é o controle do espaço aéreo, as bibliotecas também usam o Sophia. Então a gente está tendo uma Rede de Bibliotecas da Área de Aeronáutica, a Rede BIA.

*Pesquisadora:* Ah, sim, interessante também. E você mencionou que vocês já tiveram um tesouro aeroespacial, esse tesouro continua sendo utilizado, vocês ainda trabalham com ele ou vocês têm alguma outra linguagem ou vocês estão construindo? [A biblioteca desenvolve ou desenvolveu projeto construção ou manutenção/atualização de linguagem de indexação com interoperabilidade semântica?]

*Sujeito 2:* Sim, tivemos o Thesaurus Aeroespacial. Na realidade esse tesouro ele está dentro do Sophia, porque os termos todos padronizados, infelizmente não no formato de um tesouro, por que ele tinha um formato de tesouro, tinha hierarquias e a gente perdeu esse arquivo porque ele era feito ainda no IBM, o IBM gera aquelas fitas, no IBM antigo né, você gerava aquelas fitas e uma dessas fitas se perdeu e se perdeu o equipamento. E nós não julgamos como urgente a construção de um outro tesouro como era aquele porque eu tenho tudo dentro do Sophia, estou com o meu vocabulário controlado todo lá. Então o termo que entra eu já padronizo, já verifico, o termo para tá lá ele não é um termo sugerido, ele já é um termo aceito. Então o processo que a gente utilizou até para a construção do Thesaurus Aeroespacial ela mantém essa filosofia, mantenho até hoje.

*Pesquisadora:* Então eu estou entendendo que você tem uma linguagem, “a linguagem do ITA”, vamos dizer assim?

*Sujeito 2:* Isso, exatamente, e a tradução que o ITA fez dos termos, porque eu não uso os termos em inglês, eu uso em português, eu optei por traduzir. E aí é a própria estrutura, eu tô numa instituição de língua portuguesa, mas daí você fala, e o que está em inglês? Autor, título está tudo em inglês, o vocabulário está em português. E hoje em dia você consegue fazer a recuperação, hoje em dia está muito mais fácil a pesquisa, você consegue ver todos os campos, então um termo que você não tenha, ele está ali. Mas assim, os termos eles são padronizados, eu sempre tirei de algum lugar, eles são consagrados, porque na realidade qual é o critério, tem que ser um termo consagrado. O ITA, o DCTA como um todo, participa das comissões da ABNT, então alguns termos, como materiais compostos, a gente teve auxílio dos professores que participavam da comissão da ABNT de materiais compostos, então por que que é usado esse termo e não aquele termo, o próprio DCTA como um todo ele tem já essa vocação de trabalhar o termo técnico, então isso para a gente é muito natural. Ele tem que ter um sentido, tem que ter uma justificativa, ele tem que ter uma data de criação, ele tem que ter um objetivo, porque a questão do tesouro, o vocabulário controlado, dessa linguagem ela tem que estar padronizada, ela tem que estar apoiada. Por que a gente tem vários assuntos que são sinónimas, e eu sempre me lembro da professora Patrícia Marchiori, entrou lixo sai lixo, então o vocabulário é exatamente isso ele traduz todo esse trabalho, então eu tenho todo um trabalho de pesquisa, um trabalho de busca, um trabalho de significado e não prático? E você tem dentro do Sophia essa possibilidade, então eu acho que construir hierarquicamente um tesouro, com relacionamentos e outras questões, eu acho que é uma questão de tempo, ou isso com o tempo não é mais tão necessário, né. A gente tem muita preocupação com os modismos, né, e as grandes áreas, principalmente a área de computação, ela é uma beleza, né então a gente tem que ter um poder de convencimento muito forte, então Ah isso não tem tradução, então eu falo Ah mas vai ter a tradução, não tem hoje. Então a gente tem que pensar, eu ainda tenho muito termo no banho-maria, que é esse termo está aqui você aguarda por um período, no máximo seis meses, aí surgiu a tradução você vai lá e inclui. Você tem muito trabalho, você tem que estar constantemente revendo, requer do profissional que trabalha com processamento técnico estar sempre antenado, estar sempre olhando, sempre fazendo crítica dos dados. Isso é outra coisa que perguntam vocês fazem críticas dos dados? Constantemente, constantemente você tem que reavaliar. Por que você tem que manter esse fundo de informação vivo, que ele queira ser pesquisado, que seja objeto de desejo, a gente já vive uma geração que é toda digital, é só o

eletrônico, e você tem que ofertar o que tem papel também. Porque o valor é tão ou mais importante do que esse digital.

*Pesquisadora:* E esse procedimento de trabalho com os termos, você tem ele documentado?

*Sujeito 2:* O processamento técnico, ele já faz isso já é fluxo internalizado. Se você analisar do anterior, porque assim a gente teve vários *gaps*, agora a gente foi esse software do INPE até a década de 90, na década de 90 a gente foi para o SABI, que era da FURGS, mas também software livre, eles vieram e instalaram, aí já usava o IBM 4341, e aí a gente foi migrando. A biblioteca, eu não sei se eu comentei com você, mas ela tem um analista de sistemas aqui dentro. Então ela é diferenciada por isso também, ela não desenvolve os nossos programas, ela já desenvolveu, tanto que a biblioteca digital do ITA ela é quem criou, hoje ela não consegue mais atualizar, então por isso que a gente voltou para o Sophia, está voltando para o Sophia. Porque todos os dados da minha biblioteca digital, o que que ela faz, ela pega os metadados daqui e cria uma máscara aqui para consultar, nada mais justo que fique tudo dentro do Sophia, porque ele já tem essa possibilidade. Então hoje ele já está internalizado. Na época o que que a gente fazia, a gente fazia a fichinha em papel, com o termo, de qual tesouros tirou e qual o significado que tinha e sentava com o pesquisador para validar o termo. Isso é um sistema, a gente foi validando os termos assim. Então, quando ele surgiu, quando ele surgiu no tesouro, os termos que eu criei, por que são específicos da minha área. A mesma coisa com os números da CDU, Mecânica do voo é um número que não existe, o ITA criou um número para mecânica do voo, por que esse é um assunto importante e aquele acervo precisa estar separado, o meu usuário, por que a biblioteca é de livre acesso, tem que entrar no corretor da Mecânica do voo e se achar, ele não achou aquele autor, mas ele tem 500 para ela procurar, então a gente criou. Sempre validado dessa maneira, então o termo que tem dúvida é o pesquisador que valida, mas já é automático, ele não assina nenhum documento mais, entendeu. Porque nós tínhamos essa relação, chamava especialistas da área, sentava e a gente fazia a revisão. Esses termos estão consagrados, gente pode passar, estes termos estão consagrados a gente passa para a fase seguinte. Ele tem um processo, tudo que a gente faz tem uma cadeia de comando, tem um processo. É um número novo tenho que criar, ele é criado a partir e é anotado. É um índice novo, é um termo novo, ele segue o processo. Ele tem um fluxo.

*Pesquisadora:* Sim, sim. E aí seguindo aqui, eu iria te perguntar sobre os metadados, o formato que vocês estão usando para registrar a linguagem, mas sabendo que é o Sophia imagino que seja o MARC 21 autoridades?

*Sujeito 2:* E o AACR também, você tem que lembrar que a gente veio do AACR, o MARC era uma coisa nova. É o que eu sempre digo a gente não pode esquecer o legado né. na realidade a gente começou a trabalhar antes de começar a trabalhar com o MARC a gente trabalhava parametrizado, ou seja, parâmetros, e os campos tinham um número, seria o embrião do que a gente tem hoje no MARC com todos os campos, com a numeração, o que significa cada um e assim por diante.

*Pesquisadora:* E para a construção do termo mesmo as suas linguagens base são pré-coordenadas e pós-coordenadas?

*Sujeito 2:* Não, para os termos de assuntos só pós-coordenadas, pré-coordenação apenas na classificação com a CDU. Usamos tesouros, o Tesouro da NASA, a LCSH já usei mais no passado.

*Pesquisadora:* Houve dificuldades? Como a questão da coordenação está sendo resolvida?

*Sujeito 2:* Olha, até agora, eu nunca soube. A gente nunca teve essa, digamos assim, dificuldade com isso. Eu acho que isso foi crescendo naturalmente. Nós tínhamos, se você me perguntar, a gente tinha um catálogo topográfico, que eu acho que ninguém mais faz, que é justamente para bibliotecas de acervo aberto, que é a sequência em fichas dos livros na estante, de uma maneira muito simples seria a tradução. E ele já tinha uma pista maravilhosa, porque as secundárias de autor, as secundárias de título, as secundárias de autoridade e as secundárias de assunto. Então, basicamente, ao fazer aquilo, registrar mesmo que fosse em papel, você já tinha uma série de preocupações, onde você já ia construindo aquele conhecimento. Então, está muito claro, algumas coisas a gente já definiu, olha se o livro é de engenharia aeronáutica, quais são os termos que tem que vir junto na hora dessa busca, porque no início eles diziam assim, a engenharia aeronáutica não, mas a aeroespacial, mas eu também preciso ter um termo de engenharia aeronáutica, então foram coisas já consagradas no início. E a gente brinca né, que as próximas gerações já começam a fazer por osmose. Aí vai entendendo o processo, se é estatística tem que ter um descritor de economia também, um assunto né a gente falava descritor, hoje em dia palavras-chave, né, mas o que é que tem que vir junto, o que é que faz parte deste pacote, cientificamente, como é que ele tem que ser construído para fazer sentido, mas isso já era usado quando não era sistema automatizado. Então isso veio naturalmente. Se você perguntar, quais foram os processos que nós fizemos novos? O computador, só. Os princípios, eles são os mesmos e se eles são bem construídos a tendência é a aprimorar, então tudo o que a biblioteconomia trouxe de estudos só agregou, então o olhar, o trabalhar a



linguagem, trabalhar isso e aquilo, hoje em dia a gente tem web semântica, coisa que não se pensava quando começou a automação de bibliotecas, você não analisava que a gente fosse chegar no mundo que chegou hoje. Hoje você tem o pessoal de inteligência artificial do ITA pensando em mecanismos de buscas de uma linguagem natural, que já esses dados venham e aí eles começam a discutir metadados conosco e ficam surpresos de saber que aquilo que para eles é dúvida, para nós já é consagrado. Então a gente está voltando, porque as ferramentas estão evoluindo, está voltando para o pessoal que faz toda essa teoria. É muito interessante, muito interessante, eu acho que as nossas preocupações hoje realmente tem que ser o tal dos *blockchains*, como eles chamam, construir objetos digitais que não sejam corrompidos, que não possam ser alterados, em que você não tenha nenhuma modificação do texto, trabalhar dados de pesquisa, todo mundo está agora a grande coqueluche vamos depositar dados de pesquisa, e você vai ler e pensa você tem certeza o que você quer depositar isso? Qual é a equipe técnica o que vai dizer que esse dado pode ser depositado. Então hoje em dia a nossa função, o nosso olhar, ele começa a mudar, porque a parte bibliográfica de alguma maneira ela está sedimentada. Vocabulário controlado para mim, se ele é bem construído desde o início ele não é problema, ele é solução. Você consegue entender o relacionamento, a hierarquia, os termos, a sinonímia, as referências, use isto no lugar daquilo, então você consegue montar e você tem softwares dos mais simples, então se você pegar um PHL, até o PHL você consegue construir. Eu digo isso porque eu já uma vez fui convidada para uma consultoria de uma escola, e eu disse eu não tenho tempo, mas eu gosto de desafios. Mexer com o livro infantil foi um outro universo, e aí eu vi que o PHL é fantástico, ele se presta para algumas coisas. Eu fui olhar e vi nossa ele pensou nisso, ele pensou naquilo. De alguma maneira, as ferramentas elas nos ajudam. Se me permite, e eu falo se eu não tiver sendo presunçosa, eu acho que precisaria a gente ter uma interação melhor com quem faz a ferramenta, não é. Mas uma interação mais humilde porque eu acho que a comunidade sairia ganhando muito. Só que eu acho que ambos os lados precisam de uma humildade muito grande na hora que a gente vai pensar nesses dados. Eu sempre costumo dizer quando eu me reúno com a equipe eu falo “o usuário tem sempre razão” e para ele ter razão a gente tem que estar no lugar dele enquanto usuário, a gente não está fazendo para o bibliotecário, a gente está fazendo para aquele que não entende nada, mas que é um especialista na área de trabalho dele. Então você vê assim trabalhos fantásticos que eles têm aqui com fibra de carbono, para coração, então os subprodutos da pesquisa que são produtos em outras áreas do conhecimento, se a gente tem é por que alguém entrou, eu sei que eu tenho esse relatório em algum lugar porque alguém da década de 80 entrou o metadado, então tem o que recuperar. Então tudo é essa questão de a gente olhar para a pesquisa como usuário dela,

porque eu acho que fica mais fácil a gente pensar em linguagem documentária, pós-coordenado, pré-coordenado, para mim o olhar tem que ser o olhar do usuário, que ele ache tudo no mesmo lugar, que ele não fique navegando.

*Pesquisadora:* Sobre a especificidade das linguagens que você utiliza, têm dificuldade?

*Sujeito 2:* Não, o ITA ele é muito específico na área do conhecimento, eu tenho Engenharia Aeronáutica, Engenharia Mecânica e Aeronáutica, Engenharia Aeroespacial Engenharia da Computação, Engenharia Civil, Engenharia Eletrônica, então é muito específico, é muito específico. Quando você recebe um colega novo eu acho que no primeiro dia ele tem vontade de sair correndo, embora a gente seja muito cordial muito simpático, a gente teve uma colega que passou no concurso de 2014 e veio, então a gente começa dando as normas de ação, ela passa por todas as seções, mas ela mesmo se confessou que ela levou dois anos para entender o nosso idioma pronto porque a gente fala siglas, mas é só rotina. E todo o novo colega ele é tutelado, então a gente tem o tempo para sentar com ele para explicar, é tudo muito específico. Se você tem uma grande dificuldade é nos dois primeiros anos, porque são termos técnicos diferentes, áreas diferentes e você vem com uma bagagem técnica que ela é boa, mas você tem que escrever a sua história aqui. Então você tem que aprender, bom eu saí com essa bagagem com essa formação, o que que isso pode me ajudar aqui? Como é que eu pego isso aqui e aplico aqui? Tudo tem aplicação.

*Pesquisadora:* Ah, sim.

*Sujeito 2:* Eu penso que não faz sentido você ter a biblioteca mais organizada do mundo, se ela não é utilizada. Eu fiz um curso uma vez e eu ouvi isso de um pesquisador “Informação fora do tempo não serve pra nada”. E, obviamente, se a gente não se coloca no lugar dele, A gente não vai poder se antecipar às demandas dele. Então aquele olhar, olha a gente está trabalhando nessa pesquisa, a gente tem tudo, antigamente a gente já fazia isso no mundo mecânico, manual, agora no mundo eletrônico é muito mais fácil, então eu tenho que olhar as minhas competências, trazer essas competências para os softwares que eu tenho hoje, e todos, minimamente, por isso que eu te digo até o mais simples ele tem minimamente um perfil que eu possa preencher, eu posso linkar esse perfil as novas publicações que eu estou recebendo. Eu tenho algumas posturas o que eu estou percebendo que nem todo mundo compartilha dela, eu acho que tem que ser um banco de dados único, já chega você ter outras informações fragmentadas, por isso que a gente tem tudo. Já fui vaiada porque repositório institucional é repositório institucional, eu falei o meu chama biblioteca da produção, o meu não, do ITA, nosso do ITA. Eu tenho uma

informação que não tem objeto digital e tem uma informação que já nasceu e já tem objeto digital, nasceu como objeto digital. Mas está tudo em um único lugar, eu não crio esses sistemas diferentes, mas eu até entendo que esses sistemas, dependendo da complexidade da instituição, tenha que ser assim. Eu preciso analisar e entender, a gente sabe disso a gente que está aqui na biblioteca sabe que é um ambiente privilegiado, porque o número de usuários, o número de pesquisadores nos permite essa possibilidade. Estamos todos juntos, se uma sessão tem dificuldade com um termo, porque cada biblioteca tem quatro seções, todo mundo se reúne, todo mundo ajuda todo mundo senta. Então, se o processamento técnico tem uma demanda a gente vai para o processamento técnico, se a referência tem, a gente vai para a referência, a prioridade sempre vai ser a referência, porque ela que está prestando auxílio ao usuário. Então é isso.

*Pesquisadora:*

*Sujeito 2:* Tem listagem de descritores, a gente tinha a listagem. Então se você me perguntar, foi feito em fichas, mas a gente gerou uma listagem, é essa listagem que a gente faz a checagem. E a biblioteca gerencia o Sophia dentro da rede de bibliotecas do DCTA, então qualquer dúvida que eles tenham, qualquer questão a gente senta e define um procedimento padrão.

### APÊNDICE C - TRANSCRIÇÃO DE ENTREVISTA III

**Local:** Biblioteca EFLCH – Unifesp campus Guarulhos

**Duração:** 36:44 min

*Pesquisadora:* 1. Identificação da biblioteca e respondente:

*Sujeito 3:* William José Sobral, bibliotecário do processamento técnico da Biblioteca da EFLCH, Unifesp Campus Guarulhos desde 2008.

*Sujeito 4:* Emerson Issa Kamiya, bibliotecário do processamento técnico da Biblioteca da EFLCH, Unifesp campus Guarulhos desde 2011.

*Pesquisadora:* Quantas pessoas trabalham com vocês na equipe?

*Sujeito 4:* A equipe do processamento técnico da Biblioteca conta conosco, 2 bibliotecários que fazemos a catalogação e juntamente nós temos 6 estagiários que auxiliam fazendo uma pré-catalogação, por conta do grande volume de doações recebidas. Há outros 4 bibliotecários em outras funções, mas que eventualmente fazem catalogação a depender da necessidade e da possibilidade.

*Sujeito 3:* Eles fazem uma pré-catalogação, os registros ficam no sistema, mas sem os assuntos e número de classificação.

*Pesquisadora:* 3. A biblioteca desenvolve ou desenvolveu projeto construção ou manutenção/atualização de linguagem de indexação com interoperabilidade semântica? Discorra.

*Sujeito 4:* Sim, existe um projeto, mas está mais no papel, seria um projeto mais individual do bibliotecário-chefe. Os registros de autoridade são compartilhados no catálogo coletivo com as demais bibliotecas da Unifesp. Foi feita uma reunião coletiva das bibliotecas para discutir sobre uso de linguagem, qual linguagem ou linguagens usar para não ficar com o catálogo desorganizado. Foi decidido que a primeira linguagem seria a BN. Em Guarulhos usamos exclusivamente a BN, em SP e Diadema, campus com ênfase nas áreas de saúde, utilizam o DeCS (para termos da área saúde) e a BN (para termos mais genéricos), foi necessário definir isso em reunião, porque já houve caso de o bibliotecário usar o termo do DeCS de áreas mais genéricas e o conflitar com a BN. Temos problemas de comunicação entre as unidades o que prejudica a realização de adaptações ou mudanças nos vocabulários. Quando não localizamos na BN vamos a LCSH, muito por autores estrangeiros, assunto de pessoa e fazemos a tradução

e adequação do registro do termo, em caso de dúvidas da tradução procuramos em fontes como a Wikipédia, que até já dá umas pistas de remissivas e eventualmente também no Google. Para a área de História da Arte uma professora indicou vocabulários do MASP e Itaú cultural. Também chegamos a utilizar um meta-buscador de vocabulários [BARTOC], mas pesquisando chegou a localizar, no momento estava precisando de vocabulários sobre literatura infantil para um caso pontual.

*Pesquisadora:* 4. Qual software está sendo utilizado para desenvolver a linguagem e como têm sido a experiência com o software?

*Sujeito 3:* Por muito tempo, desde a implantação das bibliotecas o software utilizado era o PHL, em 2015 foi adquirido o software Pergamum. No PHL o controle de vocabulário era bem ruim, com a migração para o Pergamum os problemas com o controle de vocabulário ficaram mais evidentes. E estamos ainda corrigindo esses problemas que ficaram. O Pergamum atende bem, em comparação com o PHL é muito melhor, as dificuldades que aparecem são por conta dos problemas que ficaram no vocabulário.

*Sujeito 4:* No PHL o campo de assunto não tinha relação com a base de autoridades, então não conseguia colocar remissivas, era um campo solto. E quando migramos para o Pergamum o campo de assunto veio para um campo de assunto local, 697. E agora o Pergamum é muito mais rico os campos 600 conversam com a base de autoridades. E como somos só nós dois para fazer isso, acabamos, conforme a catalogação na busca do termo para indexar, quando percebemos problemas nos termos vamos limpando.

*Sujeito 3:* Na migração grande parte do acervo que estava catalogado e tinha os assuntos no PHL, os assuntos foram para esse campo 697, que hoje na busca por assunto no catálogo não recupera, só recupera se for pesquisa geral, então a dificuldade hoje é mais por conta disso.

*Pesquisadora:* 5. Qual formato de metadados foi adotado para desenvolver a linguagem?

*Sujeito 3:* MARC 21 Autoridades.

*Pesquisadora:* 6. Está trabalhando com linguagens que possuem modos de coordenação diferentes?

*Sujeito 3:* Sim.

*Pesquisadora:* 7. Houve dificuldades? Como a questão da coordenação está sendo resolvida?

*Sujeito 4:* Quando o termo está pré-coordenado na BN é mantido, especialmente por causa das remissivas, não é favorável "é uma faca de dois gumes". Mas, se precisar importar da BN um termo pré-coordenado ele será mantido.

*Pesquisadora:* 8. Está trabalhando com linguagens que possuem níveis de especificidade diferentes? Ex. Geral: TBN, LCSH e Específicas: MeSH, DeCS, Tesouros de áreas, Agrovoc etc.

*Sujeito 3:* Sim.

*Pesquisadora:* Houve dificuldades? Como a questão de especificidade das linguagens fonte está sendo resolvida?

*Sujeito 4:* Existe o caso das bibliotecas das unidades da área de saúde, em que são usados os termos específicos de saúde mantendo como está no DeCS e nos casos de termos genéricos, utilizar somente da BN, para não ter diferenças de hierarquia ou de significado. Vemos muita dificuldade em incorporar outras linguagens, acabando sendo atendidos pela BN, em geral. Por mais críticas à BN/LCSH, é um vocabulário utilizado no mundo todo, por exemplo, na OCLC já tem algumas indicações de termos e na prática nos atende. Mesmo vendo algumas lacunas, evitamos utilizar outros vocabulários, pelo receio de encavalar termos de vocabulários diferentes. Acabamos mais localizando termos nas outras linguagens, como no Itaú Cultural ou no MASP, e depois buscamos um correspondente na BN para tentar manter uma consistência. E em último caso criamos o termo.

*Pesquisadora:* Qual outra dificuldade foi encontrada durante o processo de desenvolvimento/atualização da linguagem com relação à interoperabilidade semântica?

*Sujeito 4:* Nossa experiência é muito na prática, com a nossa bagagem da faculdade.

*Sujeito 3:* É tudo bem na raça, tivemos que encarar o desafio do processamento técnico aqui, bem por conta de termos uma equipe pequena e grande quantidade de itens a ser catalogados, junto com as cobranças de resultados, ficamos muito limitados a resolver os problemas primeiro para depois se preocupar com outras questões. O coordenador da biblioteca tem interesse em criar um vocabulário nosso, de Humanas para o campus, ele gostaria de começar a pensar isso ainda esse ano. Há uma bibliotecária nova que entrou no ano passado para trabalhar no setor de referência, que estava precisando desenvolver, estávamos nos preocupando com outras questões, e ela detectou a falta que o vocabulário faz.

## APÊNDICE D – QUADRO DE ANÁLISE DAS TRANSCRIÇÕES - ETAPA DE PRÉ-ANÁLISE

1	Linguagem de indexação
Biblioteca A	<p><i>“Na verdade, a biblioteca separadamente não, né, a gente como faz parte da rede de bibliotecas Unesp então esse projeto é desenvolvido de forma coletiva”</i> (Biblioteca A, Sujeito 1).</p> <p><i>“[...] a gente trabalha como a linguagem integrada dentro do Aleph, o que é o software que a gente usa para fazer catalogação”</i> (Biblioteca A, Sujeito 1).</p>
Biblioteca B	<p><i>“Então eu tenho manual dessa época, com as tradições que nós fizemos, os ajustes, nós usávamos, sempre usamos, os cabeçalhos de assunto da biblioteca do Congresso de Washington, Tesouros da NASA, Thesaurus do Engineering Joint Council, usamos também o do IBICT, que ele tem um tesouro em português que foi feito em parceria seria com Portugal, vários tesouros para compor. A gente chegou a ter um Thesaurus Aeroespacial dos termos específicos da nossa área, a gente foi construindo a questão dos metadados, dos campos, a gente foi construindo com base, mas em alguma coisa muito sólida que já tinha”</i> (Biblioteca B, Sujeito 2).</p> <p><i>“É o Sophia, mas o nosso software é customizado, então tem muita gente que vem aqui vê o nosso Sophia e fala o meu não tem isso, o meu não tem aquilo e eu falo é porque é o nosso. E o Sophia foi um software eleito mediante aquela tabela de avaliação de softwares da Lígia Café, que era o instrumento que nós tínhamos a época. Então nós analisamos cinco pacotes de softwares de bibliotecas e o que mais atingiu os nossos objetivos foi o Sophia, que assim mesmo teve que ser customizado. [...] Então a avaliação do Sophia que a gente fez usando a ferramenta da Lígia Café eu acho que isso é importante. É um trabalho que foi um divisor de águas [...]. A biblioteca do ITA, o DCTA como um todo ele é muito peculiar, então é a maior biblioteca, é o maior acervo, é o maior tudo e a gente saiu na frente e nós conseguimos convencê-los com esse documento de usar um software único para todos. Então o que se precisa, aí me dizem, você prefere esse ou aquele software, eu falo não, nós não preferimos, não a gente insistiu naquele que respondeu as nossas expectativas técnicas. Então se tiver um outro no futuro, não tem problema nenhum para migrar desse para o outro, mas eles entenderam que a gente tem que ter um software único, então o CENDOC está comprando para as outras unidades”</i> (Biblioteca B, Sujeito 2).</p> <p><i>“Sim, tivemos o Thesaurus Aeroespacial. Na realidade esse tesouro ele está dentro do Sophia, porque os termos todos padronizados, infelizmente não no formato de um tesouro, porque ele tinha um formato de tesouro, tinha hierarquias e a gente perdeu esse arquivo [...]. E nós não julgamos como urgente a construção de um outro tesouro como era aquele porque eu tenho tudo dentro do Sophia, estou com o meu vocabulário controlado todo lá. Então o termo que entra eu já padronizo, já verifico, o termo para tá lá ele não é um termo sugerido, ele já é um termo aceito. Então o processo que a gente utilizou até para a construção do Thesaurus Aeroespacial ela mantém essa filosofia, mantenho até hoje”</i> (Biblioteca B, Sujeito 2).</p> <p><i>“Isso, exatamente, e a tradução que o ITA fez dos termos, porque eu não uso os termos em inglês, eu uso em português, eu optei por traduzir. [...]E hoje em dia você consegue fazer a recuperação, hoje em dia está muito mais fácil a pesquisa, você consegue ver todos os campos, então um termo que você não tenha, ele está ali. Mas assim, os termos eles são padronizados, eu sempre tirei de algum lugar, eles são consagrados, porque na realidade qual é o critério, tem que ser um termo consagrado. O ITA, o DCTA como um todo, participa das comissões da ABNT, então alguns termos, como materiais compostos, a gente teve auxílio dos professores que participavam da comissão da ABNT de materiais compostos, então por que que é usado esse termo e não aquele termo, o próprio DCTA como um todo ele tem já essa</i></p>

	<p>vocação de trabalhar o termo técnico, então isso para a gente é muito natural. Ele tem que ter um sentido, tem que ter uma justificativa, ele tem que ter uma data de criação, ele tem que ter um objetivo, porque a questão do tesouro, o vocabulário controlado, dessa linguagem ela tem que estar padronizada, ela tem que estar apoiada. Por que a gente tem vários assuntos que são sinonímias, e eu sempre me lembro da professora Patrícia Marchiori, entrou lixo sai lixo, então o vocabulário é exatamente isso ele traduz todo esse trabalho, então eu tenho todo um trabalho de pesquisa, um trabalho de busca, um trabalho de significado e não prático? E você tem dentro do Sophia essa possibilidade, então eu acho que construir hierarquicamente um tesouro, com relacionamentos e outras questões, eu acho que é uma questão de tempo, ou isso com o tempo não é mais tão necessário, né. [...] Então a gente tem que pensar, eu ainda tenho muito termo no banho-maria, que é esse termo está aqui você aguarda por um período, no máximo seis meses, aí surgiu a tradução você vai lá e inclui.” (Biblioteca B, Sujeito 2).</p> <p>“Na época o que que a gente fazia, a gente fazia a fichinha em papel, com o termo, de qual tesouros tirou e qual o significado que tinha e sentava com o pesquisador para validar o termo. Isso é um sistema, a gente foi validando os termos assim. Então, quando ele surgiu, quando ele surgiu no tesouro, os termos que eu criei, por que são específicos da minha área. [...] Sempre validado dessa maneira, então o termo que tem dúvida é o pesquisador que valida, mas já é automático, ele não assina nenhum documento mais, entendeu. Porque nós tínhamos essa relação, chamava especialistas da área, sentava e a gente fazia a revisão. Esses termos estão consagrados, gente pode passar, estes termos estão consagrados a gente passa para a fase seguinte” (Biblioteca B, Sujeito 2).</p> <p>“Vocabulário controlado para mim, se ele é bem construído desde o início ele não é problema, ele é solução. Você consegue entender o relacionamento, a hierarquia, os termos, a sinonímia, as referências, use isto no lugar daquilo, então você consegue montar e você tem softwares dos mais simples, então se você pegar um PHL, até o PHL você consegue construir” (Biblioteca B, Sujeito 2).</p>
Biblioteca C	<p>“Sim, existe um projeto [para a construção de uma linguagem de indexação], mas está mais no papel, seria um projeto mais individual do bibliotecário-chefe” (Biblioteca C, Sujeito 4).</p> <p>“Por muito tempo, desde a implantação das bibliotecas o software utilizado era o PHL, em 2015 foi adquirido o software Pergamum. No PHL o controle de vocabulário era bem ruim, com a migração para o Pergamum os problemas com o controle de vocabulário ficaram mais evidentes. E estamos ainda corrigindo esses problemas que ficaram. O Pergamum atende bem, em comparação com o PHL é muito melhor, as dificuldades que aparecem são por conta dos problemas que ficaram no vocabulário” (Biblioteca C, Sujeito 3).</p> <p>“No PHL o campo de assunto não tinha relação com a base de autoridades, então não conseguia colocar remissivas, era um campo solto. E quando migramos para o Pergamum o campo de assunto veio para um campo de assunto local, 697. E agora o Pergamum é muito mais rico os campos 600 conversam com a base de autoridades. E como somos só nós dois para fazer isso, acabamos, conforme a catalogação na busca do termo para indexar, quando percebemos problemas nos termos vamos limpando” (Biblioteca C, Sujeito 4).</p> <p>“Na migração grande parte do acervo que estava catalogado e tinha os assuntos no PHL, os assuntos foram para esse campo 697, que hoje na busca por assunto no catálogo não recupera, só recupera se for pesquisa geral, então a dificuldade hoje é mais por conta disso” (Biblioteca C, Sujeito 3).</p> <p>“O coordenador da biblioteca tem interesse em criar um vocabulário nosso, de Humanas para o campus, ele gostaria de começar a pensar isso ainda esse ano. Há uma bibliotecária nova que entrou no ano passado para trabalhar no setor de</p>



	<i>referência, que estava precisando desenvolver, estávamos nos preocupando com outras questões, e ela detectou a falta que o vocabulário faz.” (Biblioteca C, Sujeito 3).</i>
2	Formatos de metadados padrão/Protocolos computacionais
Biblioteca A	<p><i>“Ah, a gente usa o mesmo formato de metadados que a gente usa para catalogação dos livros, mas claro usando os campos específicos do MARC no que diz respeito a catalogação de assunto ou de autor. [...] É, MARC Autoridades.” (Biblioteca A, Sujeito 1).</i></p> <p><i>“As vezes, por exemplo, uma das coisas que a gente achou bem assim diferente, foi quando foi criar os cabeçalhos para personagens fictícios, quando você usa o nome da pessoa seguido do personagem fictício, por exemplo, Robson Crusóé aí entre parênteses personagem fictício. A BN tratava ele num campo e a LC em outro completamente diferente, um tratava como assunto tópico e a outra tratava como nome pessoal. E aí era uma diferença que no começo a gente ficou “Como que a gente faz?” e aí quem acabou decidindo foi o grupo né, que trata de indexação, e aí fizeram reuniões... e aí decidiram como é, para ficar padronizado né, por que às vezes a pessoa, uma de nós da rede importava da BN e ficava num determinado campo e quem importava da LC ficava em outro e era a mesma coisa. [...] O entendimento que tem de se aquela pessoa, o fictício, ele é só um assunto tópico ou se, apesar de ser fictício, ele é uma pessoa sim e deveria ser tratado como um campo 100” (Biblioteca A, Sujeito 1).</i></p>
Biblioteca B	<p><i>“Os sistemas hoje eles têm campo para tudo, tem MARC, tem tudo. Antigamente não era assim. O que a gente tinha era uma norma ISO que diz dado em rede de acervo, de informação bibliográfica, tem que ser passado dessa maneira, que é o que permite que os bancos se juntem e a gente consiga pesquisar” (Biblioteca B, Sujeito 2).</i></p> <p><i>“A biblioteca, eu não sei se eu comentei com você, mas ela tem um analista de sistemas aqui dentro. Então ela é diferenciada por isso também, ela não desenvolve os nossos programas, ela já desenvolveu, tanto que a biblioteca digital do ITA ela é quem criou, hoje ela não consegue mais atualizar, então por isso que a gente voltou para o Sophia, está voltando para o Sophia. Porque todos os dados da minha biblioteca digital, o que que ela faz, ela pega os metadados daqui e cria uma máscara aqui para consultar, nada mais justo que fique tudo dentro do Sophia, porque ele já tem essa possibilidade” (Biblioteca B, Sujeito 2).</i></p> <p><i>“E o AACR também, você tem que lembrar que a gente veio do AACR, o MARC era uma coisa nova. É o que eu sempre digo a gente não pode esquecer o legado né. na realidade a gente começou a trabalhar antes de começar a trabalhar com o MARC a gente trabalhava parametrizado, ou seja, parâmetros, e os campos tinham um número, seria o embrião do que a gente tem hoje no MARC com todos os campos, com a numeração, o que significa cada um e assim por diante” (Biblioteca B, Sujeito 2).</i></p> <p><i>“Eu tenho uma informação que não tem objeto digital e tem uma informação que já nasceu e já tem objeto digital, nasceu como objeto digital. Mas está tudo em um único lugar, eu não crio esses sistemas diferentes, mas eu até entendo que esses sistemas, dependendo da complexidade da instituição, tenha que ser assim” (Biblioteca B, Sujeito 2).</i></p>
Biblioteca C	<i>“MARC 21 Autoridades” (Biblioteca C, Sujeito 3).</i>

3	Mapeamentos semânticos
---	------------------------

Biblioteca A	<p><i>“A questão é que, na hora de fazer os cabeçalhos, na hora de pensar nessa construção, a gente acaba usando essas que são pós-coordenadas, que é a LC que é BN, que você tem essa liberdade para construir os cabeçalhos. A CDD a gente usa para identificação do material na estante, para composição do número na etiqueta, para facilitar para o aluno encontrar, para a gente conseguir ordenar o material e, a linguagem própria, a linguagem documentária a gente usa para conseguir expressar o conteúdo do documento da forma mais eficaz possível” (Biblioteca A, Sujeito 1).</i></p> <p><i>“Quando a gente precisa fazer buscas externas, sejam na LC ou na BN, a gente consegue encontrar mais os termos que a gente precisa, pois são termos assim mais consolidados, e às vezes, são bem específicos [...] são termos mais consolidados. São raras as exceções, eu acho que até hoje a gente precisou uma única vez solicitar a criação de um termo que não tinha nem na LC nem na BN e aí a gente passou por todo aquele processo de preencher um formulário, de pedir ajuda ao docente específico da área, de levantar fontes, mas foi uma única vez, nas outras vezes que a gente precisou de assuntos eles estavam contemplados ou na BN ou na própria LC, se já não estavam na própria linguagem da UNESP” (Biblioteca A, Sujeito 1).</i></p> <p><i>“E é claro como a gente usa de linguagens diferentes, inclusive, além do que eu, aqui a gente só usa a BN e a LC, mas eu vejo nas instruções que vêm, por exemplo, o pessoal de saúde usa outras linguagens específicas, então é mais só a gente aparar essas pequenas arestas, mas acaba se tornando uma linguagem mais completa possível, pelo menos no meu ponto de vista” (Biblioteca A, Sujeito 1).</i></p>
Biblioteca B	<p><i>“Não, para os termos de assuntos só pós-coordenadas, pré-coordenação apenas na classificação com a CDU. Usamos tesouros, o Tesouro da NASA, a LCSH já usei mais no passado” (Biblioteca B, Sujeito 2).</i></p> <p><i>“Olha, até agora, eu nunca soube. A gente nunca teve essa, digamos assim, dificuldade com isso. [...] Então, está muito claro, algumas coisas a gente já definiu, olha se o livro é de engenharia aeronáutica, quais são os termos que tem que vir junto na hora dessa busca, porque no início eles diziam assim, a engenharia aeronáutica não, mas a aeroespacial, mas eu também preciso ter um termo de engenharia aeronáutica, então foram coisas já consagradas no início. [...], mas o que é que tem que vir junto, o que é que faz parte deste pacote, cientificamente, como é que ele tem que ser construído para fazer sentido, mas isso já era usado quando não era sistema automatizado. Então isso veio naturalmente” (Biblioteca B, Sujeito 2).</i></p> <p><i>“Não, o ITA ele é muito específico na área do conhecimento, eu tenho Engenharia Aeronáutica, Engenharia Mecânica e Aeronáutica, Engenharia Aeroespacial Engenharia da Computação, Engenharia Civil, Engenharia Eletrônica, então é muito específico, é muito específico” (Biblioteca B, Sujeito 2).</i></p> <p><i>“Eu tenho algumas posturas o que eu estou percebendo que nem todo mundo compartilha dela, eu acho que tem que ser um banco de dados único, já chega você ter outras informações fragmentadas, por isso que a gente tem tudo. Já fui vaiada porque repositório institucional é repositório institucional, eu falei o meu chama biblioteca da produção [...]” (Biblioteca B, Sujeito 2).</i></p>

Biblioteca C	<p><i>“Os registros de autoridade são compartilhados no catálogo coletivo com as demais bibliotecas da Unifesp. Foi feita uma reunião coletiva das bibliotecas para discutir sobre uso de linguagem, qual linguagem ou linguagens usar para não ficar com o catálogo desorganizado. Foi decidido que a primeira linguagem seria a BN. Em Guarulhos usamos exclusivamente a BN, em SP e Diadema, campus com ênfase nas áreas de saúde, utilizam o DeCS (para termos da área saúde) e a BN (para termos mais genéricos), foi necessário definir isso em reunião, porque já houve caso de o bibliotecário usar o termo do DeCS de áreas mais genéricas e o conflitar com a BN. Temos problemas de comunicação entre as unidades o que prejudica a realização de adaptações ou mudanças nos vocabulários. Quando não localizamos na BN vamos a LCSH, muito por autores estrangeiros, assunto de pessoa e fazemos a tradução e adequação do registro do termo, em caso de dúvidas da tradução procuramos em fontes como a Wikipédia, que até já dá umas pistas de remissivas e eventualmente também no Google. Para a área de História da Arte uma professora indicou vocabulários do MASP e Itaú cultural. Também chegamos a utilizar um metabuscador de vocabulários [BARTOC], mas pesquisando chegou a localizar, no momento estava precisando de vocabulários sobre literatura infantil para um caso pontual”</i> (Biblioteca C, Sujeito 4).</p> <p><i>“Quando o termo está pré-coordenado na BN é mantido, especialmente por causa das remissivas, não é favorável “é uma faca de dois gumes”. Mas, se precisar importar da BN um termo pré-coordenado ele será mantido”</i> (Biblioteca C, Sujeito 4).</p> <p><i>“Existe o caso das bibliotecas das unidades da área de saúde, em que são usados os termos específicos de saúde mantendo como está no DeCS e nos casos de termos genéricos, utilizar somente da BN, para não ter diferenças de hierarquia ou de significado. Vemos muita dificuldade em incorporar outras linguagens, acabando sendo atendidos pela BN, em geral. Por mais críticas à BN/LCSH, é um vocabulário utilizado no mundo todo, por exemplo, na OCLC já tem algumas indicações de termos e na prática nos atende. Mesmo vendo algumas lacunas, evitamos utilizar outros vocabulários, pelo receio de encavalhar termos de vocabulários diferentes. Acabamos mais localizando termos nas outras linguagens, como no Itaú Cultural ou no MASP, e depois buscamos um correspondente na BN para tentar manter uma consistência. E em último caso criamos o termo.”</i> (Biblioteca C, Sujeito 4).</p>
--------------	---

4	Política de indexação institucional
Biblioteca A	<p><i>“Agora se é algum caso como foi esse do personagem, além de estabelecer como isso vai ser usado, a gente sempre recebe um feedback dizendo o porquê, ah a decisão foi essa então a gente faz assim e aí eu, por exemplo, quando eu recebo estagiárias todas essas informações que a gente têm, desde as que estão formalmente lá no site quanto essas que vão sendo trocadas por e-mail, que vão sendo respostas das nossas dúvidas, a gente vai passando, para deixar todo mundo com o mesmo nível de conhecimento do que tem que fazer no sistema”</i> (Biblioteca A, Sujeito 1).</p>

Biblioteca B	<p><i>“O processamento técnico, ele já faz isso já é fluxo internalizado. [...] A mesma coisa com os números da CDU, Mecânica do voo é um número que não existe, o ITA criou um número para mecânica do voo, por que esse é um assunto importante e aquele acervo precisa estar separado, o meu usuário, por que a biblioteca é de livre acesso, tem que entrar no corretor da Mecânica do voo e se achar, ele não achou aquele autor, mas ele tem 500 para ela procurar, então a gente criou. [...] Ele tem um processo, tudo que a gente faz tem uma cadeia de comando, tem um processo. É um número novo tenho que criar, ele é criado a partir e é anotado. É um índice novo, é um termo novo, ele segue o processo. Ele tem um fluxo”</i> (Biblioteca B, Sujeito 2).</p> <p><i>“Se você perguntar, quais foram os processos que nós fizemos novos? O computador, só. Os princípios, eles são os mesmos e se eles são bem construídos a tendência é a aprimorar, então tudo o que a biblioteconomia trouxe de estudos só agregou, então o olhar, o trabalhar a linguagem, trabalhar isso e aquilo, hoje em dia a gente tem web semântica, coisa que não se pensava quando começou a automação de bibliotecas, você não analisava que a gente fosse chegar no mundo que chegou hoje. Hoje você tem o pessoal de inteligência artificial do ITA pensando em mecanismos de buscas de uma linguagem natural, que já esses dados venham e aí eles começam a discutir metadados conosco e ficam surpresos de saber que aquilo que para eles é dúvida, para nós já é consagrado. Então a gente está voltando, porque as ferramentas estão evoluindo, está voltando para o pessoal que faz toda essa teoria”</i> (Biblioteca B, Sujeito 2).</p>
(Biblioteca C	<p><i>“Foi feita uma reunião coletiva das bibliotecas para discutir sobre uso de linguagem, qual linguagem ou linguagens usar para não ficar com o catálogo desorganizado. Foi decidido que a primeira linguagem seria a BN. Em Guarulhos usamos exclusivamente a BN, em SP e Diadema, campus com ênfase nas áreas de saúde, utilizam o DeCS (para termos da área saúde) e a BN (para termos mais genéricos), foi necessário definir isso em reunião, porque já houve caso de o bibliotecário usar o termo do DeCS de áreas mais genéricas e o conflitar com a BN. Temos problemas de comunicação entre as unidades o que prejudica a realização de adaptações ou mudanças nos vocabulários”</i> (Biblioteca C, Sujeito 4).</p>

5	Restrições do sistema
Biblioteca A	
Biblioteca B	
(Biblioteca C	<p><i>“É tudo bem na raça, tivemos que encarar o desafio do processamento técnico aqui, bem por conta de termos uma equipe pequena e grande quantidade de itens a ser catalogados, junto com as cobranças de resultados, ficamos muito limitados a resolver os problemas primeiro para depois se preocupar com outras questões.”</i> (Biblioteca C, Sujeito 3).</p>

**APÊNDICE E – QUADRO DE ANÁLISE DAS TRANSCRIÇÕES - ETAPA DE EXPLORAÇÃO DO MATERIAL**

<b>1</b>	<b>Equivalência entre linguagens</b>
Biblioteca A	
Biblioteca B	
Biblioteca C	<i>“Acabamos mais localizando termos nas outras linguagens, como no Itaú Cultural ou no MASP, e depois buscamos um correspondente na BN para tentar manter uma consistência. E em último caso criamos o termo”</i> (Biblioteca C, Sujeito 4).

<b>2</b>	<b>Formatos de intercâmbio</b>
Biblioteca A	<p><i>“[...] a gente trabalha como a linguagem integrada dentro do Aleph, o que é o software que a gente usa para fazer catalogação. [...] É, MARC Autoridades”</i> (Biblioteca A, Sujeito 1).</p> <p><i>“Exatamente. O entendimento que tem de se aquela pessoa, o fictício, ele é só um assunto tópico ou se, apesar de ser fictício, ele é uma pessoa sim e deveria ser tratado como um campo 100”</i> (Biblioteca A, Sujeito 1).</p>
Biblioteca B	<p><i>“A gente chegou a ter um Thesaurus Aeroespacial dos termos específicos da nossa área, a gente foi construindo a questão dos metadados, dos campos, a gente foi construindo com base, mas em alguma coisa muito sólida que já tinha. Então, nós tínhamos toda uma estrutura que precisava ser convertida agora para alguma coisa que a máquina entendesse, então nós fazemos ainda os processos à moda antiga, a gente cataloga, classifica e indexa e hoje isso tem uma única linguagem. [...] Os sistemas hoje eles têm campo para tudo, tem MARC, tem tudo. Antigamente não era assim. O que a gente tinha era uma norma ISO que diz dado em rede de acervo, de informação bibliográfica, tem que ser passado dessa maneira, que é o que permite que os bancos se juntem e a gente consiga pesquisar”</i> (Biblioteca B, Sujeito 2).</p> <p><i>“Na realidade esse tesouro ele está dentro do Sophia, porque os termos todos padronizados, infelizmente não no formato de um tesouro, por que ele tinha um formato de tesouro, tinha hierarquias e a gente perdeu esse arquivo porque ele era feito ainda no IBM, o IBM gera aquelas fitas, no IBM antigo né, você gerava aquelas fitas e uma dessas fitas se perdeu e se perdeu o equipamento. E nós não julgamos como urgente a construção de um outro tesouro como era aquele porque eu tenho tudo dentro do Sophia, estou com o meu vocabulário controlado todo lá”</i> (Biblioteca B, Sujeito 2).</p> <p><i>“Porque todos os dados da minha biblioteca digital, o que que ela faz, ela pega os metadados daqui e cria uma máscara aqui para consultar, nada mais justo que fique tudo dentro do Sophia, porque ele já tem essa possibilidade”</i> (Biblioteca B, Sujeito 2).</p>

Biblioteca C	<p><i>“Por muito tempo, desde a implantação das bibliotecas o software utilizado era o PHL, em 2015 foi adquirido o software Pergamum. No PHL o controle de vocabulário era bem ruim, com a migração para o Pergamum os problemas com o controle de vocabulário ficaram mais evidentes. E estamos ainda corrigindo esses problemas que ficaram. O Pergamum atende bem, em comparação com o PHL é muito melhor, as dificuldades que aparecem são por conta dos problemas que ficaram no vocabulário”</i> (Biblioteca C, Sujeito 3).</p> <p><i>“No PHL o campo de assunto não tinha relação com a base de autoridades, então não conseguia colocar remissivas, era um campo solto. E quando migramos para o Pergamum o campo de assunto veio para um campo de assunto local, 697. E agora o Pergamum é muito mais rico os campos 600 conversam com a base de autoridades. E como somos só nós dois para fazer isso, acabamos, conforme a catalogação na busca do termo para indexar, quando percebemos problemas nos termos vamos limpando”</i> (Biblioteca C, Sujeito 4).</p> <p><i>“Na migração grande parte do acervo que estava catalogado e tinha os assuntos no PHL, os assuntos foram para esse campo 697, que hoje na busca por assunto no catálogo não recupera, só recupera se for pesquisa geral, então a dificuldade hoje é mais por conta disso”</i> (Biblioteca C, Sujeito 3).</p> <p><i>“MARC 21 Autoridades.”</i> (Biblioteca C, Sujeito 3).</p>
--------------	--

3	Tipos de mapeamento
Biblioteca A	
Biblioteca B	
Biblioteca C	

4	Mapeamento de equivalência
Biblioteca A	
Biblioteca B	
Biblioteca C	

5	Equivalência exata, inexata e parcial
---	---------------------------------------

Biblioteca A	
Biblioteca B	
Biblioteca C	

<b>6</b>	<b>Uso de mapeamentos em recuperação da informação</b>
Biblioteca A	
Biblioteca B	
Biblioteca C	

<b>7</b>	<b>Lidando com a pré-coordenação</b>
Biblioteca A	
Biblioteca B	<p><i>“Não, para os termos de assuntos só pós-coordenadas, pré-coordenação apenas na classificação com a CDU. Usamos tesouros, o Tesouro da NASA, a LCSH já usei mais no passado” (Biblioteca B, Sujeito 2).</i></p> <p><i>“Olha, até agora, eu nunca soube. A gente nunca teve essa, digamos assim, dificuldade com isso. [...] Então, está muito claro, algumas coisas a gente já definiu, olha se o livro é de engenharia aeronáutica, quais são os termos que tem que vir junto na hora dessa busca, porque no início eles diziam assim, a engenharia aeronáutica não, mas a aeroespacial, mas eu também preciso ter um termo de engenharia aeronáutica, então foram coisas já consagradas no início. [...] se é estatística tem que ter um descritor de economia também, um assunto né a gente falava descritor, hoje em dia palavras-chave, né, mas o que é que tem que vir junto, o que é que faz parte deste pacote, cientificamente, como é que ele tem que ser construído para fazer sentido, mas isso já era usado quando não era sistema automatizado. Então isso veio naturalmente” (Biblioteca B, Sujeito 2).</i></p>
Biblioteca C	<p><i>“Quando o termo está pré-coordenado na BN é mantido, especialmente por causa das remissivas, não é favorável “é uma faca de dois gumes”. Mas, se precisar importar da BN um termo pré-coordenado ele será mantido” (Biblioteca C, Sujeito 4).</i></p>

<b>8</b>	<b>Apresentação de vocabulários mapeados</b>
Biblioteca A	
Biblioteca B	
Biblioteca C	