

UNIVERSIDADE FEDERAL DE SÃO CARLOS (UFSCar)  
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS (CECH)  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO (PPGCI)

SUELEN CAMILO FERREIRA

**MODELO SKOS E ISO 25964: SIMILARIDADES DAS  
RECOMENDAÇÕES PARA REPRESENTAÇÃO DE TESAUROS**

São Carlos

2023

SUELEN CAMILO FERREIRA

MODELO SKOS E ISO 25964: SIMILARIDADES DAS  
RECOMENDAÇÕES PARA A REPRESENTAÇÃO DE TESAUROS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos como requisito para obtenção do título de Mestre em Ciência da Informação.

**Área de concentração:** Conhecimento, Tecnologia e Inovação.

**Linha de pesquisa:** Tecnologia, Informação e Representação.

**Orientadora:** Profa. Dra. Paula Regina Dal'Evedove.

São Carlos  
2023

F383m

Ferreira, Suelen Camilo

Modelo SKOS e ISO 25964 : similaridades das recomendações para a representação de tesouros / Suelen Camilo Ferreira. – 2023.  
168f.

Dissertação (Mestrado) - Universidade Federal de São Carlos, campus São Carlos, São Carlos

Orientador (a): Paula Regina Dal'Evedove

Banca Examinadora: Brígida Maria Nogueira Cervantes, Ana Lúcia Terra, Rogério Aparecido Sá Ramalho.

Bibliografia

1. Simple Knowledge Organization System. 2. ISO 25964. 3. Tesouro.  
I. Orientador. II. Ferreira, Suelen Camilo. III. Título.

CDD: 025  
CDU: 004.93

Bibliotecária responsável: Suelen Camilo Ferreira - CRB-8/9642



**UNIVERSIDADE FEDERAL DE SÃO CARLOS**

Centro de Educação e Ciências Humanas  
Programa de Pós-Graduação em Ciência da Informação

---

**Folha de Aprovação**

---

Defesa de Dissertação de Mestrado da candidata Suelen Camilo Ferreira, realizada em 14/08/2023.

**Comissão Julgadora:**

Profa. Dra. Paula Regina Dal'Evedove (UFSCar)

Prof. Dr. Rogério Aparecido Sá Ramalho (UFSCar)

Profa. Dra. Brígida Maria Nogueira Cervantes (UEL)

Profa. Dra. Ana Lúcia Silva Terra (UC)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Ciência da Informação.

## DEDICATÓRIA

*Dedico à minha família, especialmente aos meus pais, ao meu companheiro de vida e ao meu filho. Para a minha mãe, Natalina, porque, mesmo sem ter tido a oportunidade de estudar, me ensinou a valorizar as instituições de ensino, as professoras e os professores. Para o meu pai, Eliseu, porque me ensinou a amar as palavras, os livros e o conhecimento. Para o Brayan, de quem vejo o brilho nos olhos ao me ouvir falar sobre esta pesquisa. E para o meu filho, Ian, pois sua chegada iluminou a minha vida e a reta final desta dissertação.*

## **AGRADECIMENTOS**

Ao Pai, Filho e Espírito Santo - a Trindade que sempre guia a minha vida e acolhe as minhas orações.

Agradeço ao meu companheiro de vida, Brayan, por todo amor, apoio e pelas incontáveis maneiras que encontrou para que eu pudesse me debruçar sobre este estudo.

Agradeço à minha mãe por jamais ter deixado de acreditar nesta batalha.

Ao meu pai, que sempre se dispõe a ler e a comentar tudo o que escrevo.

Ao meu filho amado que mesmo sem entender, compartilhou comigo cada momento de entusiasmo e frustração enquanto eu escrevia e o amamentava e, em seguida, tinha de entregá-lo aos prantos para o colo do pai dele.

Agradeço à minha irmã Isabelly, pois me ajudou a cuidar do lan, da casa, da comida, da roupa e da minha autoestima enquanto ninguém mais podia.

A toda a minha família, que sempre compreendeu a minha ausência sem deixar de estar presente de alguma maneira.

Agradeço às minhas terapeutas por me ajudarem a percorrer esta pesquisa sem perder o rumo da minha jornada pelo autoconhecimento.

Agradeço aos professores do Curso de Biblioteconomia e Ciência da Informação da Universidade Federal de São Carlos por todos os ensinamentos e oportunidades, indicando caminhos instigantes para uma atuação científica e profissional que perseguisse a tríade ensino-pesquisa-extensão.

Agradeço à minha orientadora Profa. Dra. Paula Regina Dal'Evedove por toda compreensão, pelos diálogos e orientações objetivas e acolhedoras, encorajando-me a prosseguir firmemente até a conclusão deste trabalho.

Agradeço a Profa. Dra. Brígida Maria Nogueira Cervantes e ao Prof. Dr. Rogério Ramalho por prontamente aceitarem compor a banca examinadora, apontando valiosas e construtivas críticas e pontos de melhoria que me deram ainda mais segurança para finalizar esta pesquisa.

A todos os docentes do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos pelas aulas, referenciais teóricos, avaliações e encorajamentos.

Agradeço aos meus primeiros ex-alunos do Curso Técnico em Biblioteconomia da Escola Técnica Estadual Parque da Juventude / Centro Paula Souza, do componente Organização, Tratamento e Recuperação da Informação, por terem me incentivado a continuar estudando enquanto me (re)apresentavam o ambiente da sala de aula como um dos meus lugares preferidos no mundo.

Agradeço à toda a equipe da biblioteca da Escola de Comunicações e Artes da Universidade de São Paulo por terem recebido a mim e ao meu bebê com total profissionalismo e humanidade, disponibilizando o acesso à norma ISO 25964, partes 1 e 2. Este apoio foi fundamental para compor os resultados do presente estudo, pois tive certa dificuldade financeira e geográfica para localizar e acessar a norma.

A todos os amigos, amigas, colegas de trabalho, companheiros de ativismo social e parceiros cujas indagações de alguma forma me induziram a alcançar este sonho.

“Eu tenho a medida que designo - este é o esplendor de se ter uma linguagem. Mas eu tenho muito mais à medida que não consigo designar. A realidade é a matéria-prima, a linguagem é o modo como vou buscá-la - e como não acho. Mas é do buscar e não achar que nasce o que eu não conhecia, e que instantaneamente reconheço. A linguagem é o meu esforço humano. Por destino tenho que ir buscar e por destino volto com as mãos vazias. Mas - volto com o indizível. O indizível só me poderá ser dado através do fracasso de minha linguagem. Só quando falha a construção, é que obtenho o que ela não conseguiu”. Clarice Lispector (1964)

## RESUMO

O modelo de dados *Simple Knowledge Organization System* (SKOS) serve para a disponibilização e compartilhamento interoperável de diferentes tipos de Sistemas de Organização do Conhecimento (SOC) no ambiente Web. Os SOC do tipo tesouros cumprem importante função na recuperação da informação e possuem a norma ISO 25964 como específica para sua construção e mapeamento. Neste cenário, dois fatores são considerados: o modelo SKOS já era amplamente utilizado quando a norma ISO 25964 surgiu; e as recomendações SKOS contemplam diversos tipos de SOC, enquanto as recomendações ISO 25964 visam tesouros. Esses fatores indicam possíveis similaridades entre as recomendações SKOS e as recomendações da ISO 25964. Mapear e analisar elementos que compõem o modelo e a norma contribui no estudo e desenvolvimento de instrumentos de representação temática como os tesouros conceituais. Neste sentido, objetiva-se apresentar similaridades entre o modelo SKOS e a norma ISO 25964 para a representação de tesouros. Trata-se de uma pesquisa exploratória e descritiva, de abordagem qualitativa, conduzida mediante pesquisa bibliográfica em bases de dados nacionais e internacionais. O *corpus* analisado totalizou 25 publicações. Foram identificadas e discutidas 11 categorias de similaridades entre o modelo SKOS e a norma ISO 25964 para construção de tesouros. Constata-se que os resultados conferem aportes teóricos e metodológicos para construção, migração, atualização e manutenção de tesouros conceituais e interoperáveis semanticamente em ambiente Web. No entanto, poucas pesquisas discutem as similaridades entre o modelo SKOS e a norma ISO 25964. São apresentadas sugestões de pesquisas futuras e complementares.

**Palavras-chave:** *Simple Knowledge Organization System*, ISO 25964, Tesouros, Sistemas de Organização do Conhecimento.

## ABSTRACT

The Simple Knowledge Organization System (SKOS) data model serves for the availability and interoperable sharing of different types of Knowledge Organization Systems (KOS) in the Web environment. Thesaurus-type KOS play an important role in information retrieval and adhere to the ISO 25964 standard specifically for their construction and mapping. In this scenario, two factors are considered: the SKOS model was already widely used when the ISO 25964 standard emerged, and SKOS recommendations encompass various types of KOS, while ISO 25964 recommendations focus on thesauri. These factors indicate potential similarities between SKOS recommendations and ISO 25964 recommendations. Mapping and analyzing elements that compose the model and the standard contribute to the study and development of thematic representation instruments such as conceptual thesauri. In this context, the objective is to present similarities between the SKOS model and the ISO 25964 standard for the representation of thesauri. This is an exploratory and descriptive research with a qualitative approach, conducted through bibliographic research in national and international databases. The analyzed corpus totaled 25 publications. Eleven categories of similarities between the SKOS model and the ISO 25964 standard for the construction of thesauri were identified and discussed. The results provide theoretical and methodological contributions for the construction, migration, updating, and semantically interoperable maintenance of conceptual thesauri in the Web environment. However, few studies discuss the similarities between the SKOS model and the ISO 25964 standard. Suggestions for future and complementary research are presented.

**Keywords:** Simple Knowledge Organization System, ISO 25964, Thesauri, Knowledge Organization Systems.

## LISTA DE FIGURAS

Figura 1 - Método quadripolar de pesquisa: interação entre os polos.	39
Figura 2 - Interface online do Tesouro de Folclore e Cultura Popular Brasileira.	67
Figura 3 - Tripla RDF.	77
Figura 4 - Relacionamento entre skos:Concept e skos:ConceptScheme.	79
Figura 5 - Modelo de dados definido pela ISO 25964.	112
Figura 6 - Correspondência entre o Tesouro da Unesco e o Eurovoc.	121

## LISTA DE QUADROS

Quadro 1 - Estrutura do relatório de pesquisa	41
Quadro 2 - Estratégias de busca de acordo com bases de dados e resultados.	43
Quadro 3 - Função de notações terminológicas de tesouros clássicos.	68
Quadro 4 - Funções das notações para conceitos e esquemas de conceitos.	79
Quadro 5 - Notações e propriedades de etiquetagem do modelo SKOS.	81
Quadro 6 - Notações de relações semânticas entre conceitos no modelo SKOS.	82
Quadro 7 - Silogismo com termo e conceito.	88
Quadro 8 - Bibliografia analisada e respectivas propostas e grupos	96
Quadro 9 - Classes na ISO 25964 e no SKOS para tesouros, conceitos e termos.	114
Quadro 10 - Relacionamentos para tesouros, conceitos e termos na norma ISO 25964 e no modelo SKOS.	115
Quadro 11 - Notação para grupos de conceitos na ISO 25964 e SKOS.	118
Quadro 12 - Notação de relacionamentos para grupos de conceitos na ISO 25964 e no SKOS.	118
Quadro 13 - Comparação de relações hierárquicas na ISO 25964 e SKOS.	125
Quadro 14 - Tipos de relação hierárquica.	127
Quadro 15 - Relacionamentos para conceitos principais na ISO 25964 e SKOS.	129
Quadro 16 - Relacionamentos associativos em ISO 25964 e SKOS.	129
Quadro 17 - Comparação entre elementos de notas da ISO 25964 e do SKOS.	130
Quadro 18 - Relacionamentos para arranjos na ISO 25964 e no modelo SKOS.	132
Quadro 19 - Elementos da extensão SKOS-XL.	138
Quadro 20 - Exemplos de elementos da extensão ISO-THES.	140
Quadro 21 - Exemplo de elementos do UNESKOS <i>Vocabulary Specification</i>	141

## LISTA DE ABREVIATURAS E SIGLAS

AAT *Art and Architecture Thesaurus*

ABNT Associação Brasileira de Normas Técnicas

AGRIS *International Information System for the Agricultural Sciences and Technology*

ANVISA Agência Nacional de Vigilância Sanitária

ASP *Academic Search Premier*

AVA Ambientes Virtuais de Aprendizagem

CESSDA *Consortium of European Social Science Data Archives*

BDTD Biblioteca Digital Brasileira de Teses e Dissertações

BENANCIB Base de Dados do Encontro Nacional de Pesquisa em Ciência da Informação

BRAPCI Base Referencial de Artigos de Periódicos em Ciência da Informação

BT *Broader Term*

BTG *Broader Term Generic*

BTGE *Broader Term Generic Extended*

BTI *Broader Term Instantial*

BTIE *Broader Term Instantial Extended*

BTP *Broader Term Partitive*

BTPE *Broader Term Partitive Extended*

CDD *Dewey Decimal Classification*

CDU Classificação Decimal Universal

CRG *Classification Research Group*

DOI *Digital Object Identifier*

ECA Escola de Comunicações e Artes

ELSST European Language Social Science Thesaurus

ESRC *Economic and Social Research Council*

FAO *Food and Agriculture Organization*

FID Federação Internacional de Documentação

FRSAD *Functional Requirements for Subject Authority Data*

HASSET Humanities and Social Science Electronic Thesaurus

HTTP *Hypertext Transfer Protocol*

HTTPS *Hyper Text Transfer Protocol Secure*

IA Inteligência Artificial  
IBICT Instituto Brasileiro de Informação em Ciência e Tecnologia  
IIB Instituto Internacional de Bibliografia  
IRI Identificador de Recurso Internacionalizado  
ISO *International Standard Organization*  
ISKO *International Society for Knowledge Organization*  
ISTA *Information Science & Technology Abstracts*  
KO *Knowledge Organization*  
KOP *Knowledge Organization Process*  
KOS *Knowledge Organization System*  
LISA *Library, Information Science & Technology Abstracts*  
LN Linguagem Natural  
LC *Library of Congress*  
LD Linguagens Documentárias  
LOD *Linked Open Data*  
LOD KOS *Linked Open Data Knowledge Organization System*  
MARC *Machine Readable Cataloging*  
MER Modelo Entidade Relacionamento  
NT *Narrower Term*  
OPAC *Online Public Access Catalog*  
ORC Organização e Representação do Conhecimento  
OWL *Ontology Web Language*  
PLN Processamento de Linguagem Natural  
PO *Product Owners*  
PPGCI Programa de Pós-Graduação em Ciência da Informação  
RDF *Resource Description Framework*  
RI Recuperação da Informação  
RT *Related Term*  
SciELO *Scientific Electronic Library Online*  
SKOS *Simple Knowledge Organization System*  
SKOS-XL *SKOS eXtension for Labels*  
SOC Sistemas de Organização do Conhecimento  
SPARQL *SPARQL Protocol and RDF Query Language*  
SRI Sistema de Recuperação da Informação

SWAD-E *Semantic Web Advanced Development for Europe*  
TA Termo Associado  
TE Termo Específico  
TEP Termo Específico Partitivo  
TEST *Thesaurus of Engineering and Scientific Terms*  
TG Termo Genérico  
TGN *Thesaurus of Geographic Names*  
TGP Termo Genérico Partitivo  
TIC Tecnologias de Informação e Comunicação  
TTI Tratamento Temático da Informação  
TR Termo Relacionado  
UE União Europeia  
UFSCar Universidade Federal de São Carlos  
ULAN *Union List of Artist Names*  
UNESCO *United Nations Educational, Scientific and Cultural Organization*  
UNESP Universidade Estadual Paulista “Júlio de Mesquita Filho”  
UP Usado Para  
URI *Uniform Resource Identifier*  
URL *Uniform Resource Locator*  
USP Universidade de São Paulo  
W3C *World Wide Web Consortium*  
XML *eXtensible Markup Language*

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>14</b>
1.1 Problematização.....	24
1.2 Justificativas e Hipóteses.....	25
1.3 Objetivos.....	30
1.4 Abordagem metodológica.....	31
1.5 Estruturação da pesquisa.....	32
<b>2 PROCEDIMENTOS METODOLÓGICOS: MODELO METODOLÓGICO QUADRIPOLAR.....</b>	<b>34</b>
2.1 Enquadramento metodológico: Modelo Metodológico Quadripolar.....	34
2.2 Descrição dos procedimentos técnicos.....	42
<b>3 ENCADEAMENTOS ENTRE CIÊNCIA DA INFORMAÇÃO E TESAUROS.....</b>	<b>48</b>
3.1 Um percurso epistemológico em Ciência da Informação.....	48
3.2 Linguagem, informação, representação temática da informação e recuperação da informação.....	55
3.3 Tesouro: instrumento terminológico, SOC e linguagem documentária.....	62
<b>4 SIMPLE KNOWLEDGE ORGANIZATION SYSTEM E NORMA ISO 25964.....</b>	<b>72</b>
4.1 Simple Knowledge Organization System.....	72
4.2 Aspectos normativos: norma ISO 25964.....	87
<b>5 ANÁLISE COMPARATIVA ENTRE A NORMA ISO 25964 E O MODELO SKOS.....</b>	<b>94</b>
5.1 Apresentação do corpus analisado.....	94
5.2 Tesouro em relação a outros SOC.....	101
5.3 Conceitos e termos.....	110
5.4 Domínios, microtesouro e grupos de conceitos.....	117
5.5 Relações de equivalência.....	120
5.6 Equivalência composta.....	122
5.7 Relações hierárquicas.....	125
5.8 Principais termos (topTerms) e principais conceitos (topConcepts).....	128
5.9 Relações associativas.....	129
5.10 Notas.....	130
5.11 Arranjos.....	131
5.12 Abordagens de aprimoramento, extensões e usos.....	132
<b>6 CONSIDERAÇÕES FINAIS.....</b>	<b>148</b>
6.1 Limitações da pesquisa.....	151
6.2 Sugestões de pesquisas futuras e complementares.....	152
<b>REFERÊNCIAS.....</b>	<b>154</b>

## 1 INTRODUÇÃO

O presente trabalho se enquadra na segunda linha de pesquisa do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos (PPGCI/UFSCar), denominada Tecnologia, Informação e Representação, com aportes às investigações relacionadas ao tema representação temática da informação em ambientes digitais.

A produção, registro, transmissão e apropriação do conhecimento sempre estiveram presentes na história da humanidade, de início, por meio de pinturas rupestres ou de ensinamentos oferecidos por meio da oralidade (Burke, 2003; Chaves, 2018). Com o surgimento das primeiras bibliotecas, gradualmente foram instituídos processos e instrumentos de tratamento do conhecimento registrado.

As bibliotecas medievais produziam e guardavam manuscritos redigidos à mão numa época em que o processo de produção e reprodução de um livro era mais lento (Burke, 2003). Mas, em meados do século XV, quando Gutenberg inventa a prensa tipográfica, são facilitadas as etapas de produção, registro e compartilhamento do conhecimento, aumentando, assim, a quantidade de informações a serem organizadas em prol do atendimento às necessidades de consultas futuras (Burke, 2003). Diante desta e de outras mudanças (como a invenção da iluminação elétrica no século XIX) emergem também necessidades específicas de lidar com um volume cada vez maior de informações, num contexto em que as ramificações e especializações técnicas e científicas ocorriam em diferentes áreas de atuação da comunidade intelectual (Burke, 2003).

Diante dos contínuos avanços da ciência, tecnologia e sociedade intensificados no início do século XX, surge com o campo da Documentação (assim cunhado por Paul Otlet e continuado por Suzanne Briet) a ideia de sistemas de informação que pudessem universalizar o conhecimento de forma qualitativa e quantitativa (Araújo, 2017; Cervantes *et al.*, 2018; Santos; Vidotti, 2009). Nesse contexto, em meio à “explosão informacional”, o surgimento da Ciência da Informação enquanto um campo se dá após a Segunda Guerra Mundial, numa época de crescimento exponencial de publicações técnicas, científicas e registros de informação de diversos tipos que se acumulavam em ritmo acelerado (Saracevic,

2009). Assim, com o passar do tempo, são desenvolvidos diversos estudos, técnicas, discussões e instrumentos no entorno dos procedimentos voltados para a organização, tratamento e recuperação da informação.

Araújo (2017), ao resgatar o que fora defendido por Capurro em 2003 no Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), aponta três paradigmas ou modelos de estudo da Ciência da Informação – físico, cognitivo e social –, que, ao serem analisados junto a treze teorias<sup>1</sup> contemporâneas da Ciência da Informação, revelam a existência de algumas inquietações sobre a forma como o conceito de informação é compreendido. Visto que as treze teorias contemporâneas se aproximam mais do paradigma social, o que demonstra que a informação existe em dimensões que não são apenas mecânicas e operacionais atreladas às dimensões física e cognitiva. A informação pode ser entendida como algo mais amplo e complexo à medida que as identidades e construção dos sujeitos sociais são construídas a partir da criação e apropriação de conhecimento, o que faz com que a informação esteja atrelada a conceitos de “cultura, memória, apropriação, mediação, documento, saberes” (Araújo, 2019, p. 29).

Os aspectos sociais que permeiam todos os processos de tratamento e uso da informação estão cada vez mais em evidência na maioria das pesquisas contemporâneas, mas também já foram constatadas no passado. Conforme resgatado por Santos e Vidotti (2009), Paul Otlet, já no início do século XX, em sua busca pela universalização do conhecimento, supunha a existência futura de uma rede global de computadores que “permitiria às pessoas procurar e pesquisar por meio de milhões de arquivos de documentos, imagens, áudio e vídeo interligados” (Santos; Vidotti, 2009, p. 3). Visionário, Paul Otlet descreveu ainda que as pessoas enviariam mensagens umas às outras por meio de aparatos que também permitiriam o compartilhamento de arquivos e agrupamento em redes sociais – “Ele denominou a coisa toda de uma ‘rede’ (*réseau*) no sentido que nós conhecemos atualmente” (Wright, 2008, p. 1 *apud* Santos; Vidotti, 2009, p. 3).

---

<sup>1</sup> As treze teorias contemporâneas apresentadas por Araújo (2017) são identificadas a partir de um amplo processo de investigação, sendo elas: 1) Análise de domínio; 2) Altimetria; 3) Cultura organizacional; 4) Curadoria digital; 5) Folksonomias e Indexação social; 6) Ética intercultural da informação; 7) Neodocumentação; 8) Humanidades digitais; 9) Arqueologia da sociedade da informação; 10) Práticas informacionais; 11) Regimes de informação; 12) Memória e; 13) aproximações com Arquivologia, Biblioteconomia e Museologia.

Para Santos e Vidotti (2009, p. 2) “compreender a tecnologia inclui a totalidade da nossa cultura material”. Haja vista que as Tecnologias de Informação e Comunicação (TIC) se constituem não apenas como ferramentas a serem investigadas e aplicadas no âmbito da Ciência da Informação, mas também, na forma como estão estabelecidas, refletem a cultura da sociedade que as utiliza e desenvolve para manusear, acessar, consumir, produzir e compartilhar informações.

O número cada vez maior de dados e informações gerados na Web, seja por meio das redes sociais ou outras plataformas digitais, demanda novos modelos explicativos da Ciência da Informação enquanto ciência que investiga teorias e práticas relacionadas à coleta, armazenamento, recuperação e uso eficaz da informação.

Para que a informação seja manipulada de forma adequada, a representação da informação é um fator fundamental na cadeia de tratamento, organização e recuperação da informação, pois é por meio da representação que os recursos informacionais poderão ser organizados e recuperados de forma ágil e relevante, seja na *Web* ou em outros ambientes informacionais. A representação da informação pode ser entendida tangencialmente ao conceito de catalogação que, desmembrado em catalogação descritiva, catalogação de assunto e classificação, tem, nestes dois últimos, relação com o tratamento temático da informação enquanto um dos processos que visa garantir recuperação informacional e acesso a itens documentais a partir das temáticas abordadas por estes (Foskett, 1973). O Tratamento Temático da Informação (TTI) consiste na elaboração de metadados de localização documental que poderão ser interpretados tanto pelo público humano como por agentes computacionais (Machado; Zafalon, 2020).

A área de TTI evoluiu ao longo do tempo, desenvolvendo-se com base em três abordagens distintas que podem ser descritas da seguinte maneira: catalogação de assuntos (com origem predominantemente norte-americana), indexação (com origem predominantemente inglesa) e análise documental (com origem predominantemente francesa) (Guimarães, 2008; 2009).

Historicamente, a abordagem da catalogação de assuntos surgiu influenciada pela Escola de Chicago e focada nas atividades bibliotecárias, enfatizando o catálogo como resultado do tratamento da informação. Essa ideia se originou

diretamente dos fundamentos da catalogação alfabética de Cutter e da tradição de usar cabeçalhos de assunto da *Library of Congress* (LC) (Guimarães, 2008; 2009).

Uma segunda abordagem na área de TTI é baseada na indexação, abrangendo bibliotecas tradicionais, centros de documentação e o universo editorial. Nessa visão, os índices são produtos do TTI que utilizam linguagens de indexação, como tesouros (Guimarães, 2008; 2009).

Na evolução da área de TTI, houve uma reação à influência norte-americana de abordagem pragmática e normalizadora, a partir das teorias de Ranganathan. Essa corrente de matriz inglesa também teve impacto significativo no Brasil (Guimarães, 2008; 2009), influenciando estudos e pesquisas realizados por grupos de pesquisadores em programas de pós-graduação em Ciência da Informação.

A abordagem da análise documental enfoca o próprio processo de TTI, com ênfase nos procedimentos de identificação e seleção de conceitos para gerar produtos. Essa linha teve forte interface com a Linguística e a Lógica. No Brasil, a partir dos anos 80, o Grupo TEMMA da Escola de Comunicações e Artes da USP, coordenado por Johanna W. Smit, desenvolveu significativamente essa concepção (Guimarães, 2008; 2009).

Nesse contexto, a Ciência da Informação é um campo interdisciplinar que fundamenta a Biblioteconomia e outras áreas profissionais. Seu universo epistemológico está centrado na produção, organização e uso da informação, etapas do ciclo de operações documentárias (Guimarães, 2009).

Diante da catalogação de assunto, indexação e análise documental - fica claro que todas elas se preocupam com o desenvolvimento de processos específicos, utilizando um conjunto de instrumentos para gerar produtos que permitam ou facilitem a recuperação da informação (Guimarães, 2009).

Para Cervantes *et al.* (2018), aliar o estudo da Representação e Recuperação da Informação a contribuições de aplicações de tecnologias da informação também culmina na representação de conceitos, por meio de termos, de determinada área do conhecimento. Estes termos compõem um vocabulário consistente “que compõe a base para a representação e a recuperação temática da informação” (Cervantes *et al.*; 2018, p. 410).

Araújo (2017, p. 16) aponta que “representação é algo socialmente construído; e é o uso feito pelas pessoas que cria os significados dos documentos”. Nesse sentido, a linguagem é um dos elementos principais para a construção de significados. E é por meio da linguagem, a partir do desenvolvimento e utilização de códigos, sintaxe e semântica específicas de diferentes culturas e domínios, que a representação da informação se torna possível. Para Gracioso (2016, p. [1]) “a linguagem está onde a informação está e onde a sociedade está” e que “a compreensão sobre a relação sujeito–mundo (mediados pela tecnologia ou não) é fenomenológica, cibernética, psicológica, ecológica e transformadora”. A linguagem pode ser entendida “como o grande instrumento potencializador do raciocínio humano” (Gracioso, 2016, p. [1]), e o entendimento da sofisticação deste instrumento como base para o surgimento de linguagens específicas – tais como: linguagem de programação, linguagem natural e linguagem documentária – corrobora com a abordagem de que, no início do século XX, a linguagem começa a ser entendida de acordo com a sua “função comunicacional, de modo que a construção de seus sentidos se estabeleceria a partir das trocas simbólicas praticadas por sujeitos participantes de uma ação de comunicação” (Gracioso; Saldanha, 2011, p. 65). Esta concepção permite pensar que essa interação comunicativa por meio da linguagem “se aproxima do plano de significação nos espaços de ações interativas virtuais” (Gracioso; Saldanha, 2011, p. 65). Nesse contexto, em abordagem à obra de Wittgenstein, “a linguagem não poderia ser entendida a partir da lógica – ela precisaria ser entendida a partir de seu uso” (Gracioso; Saldanha, 2011, p. 65).

Considerando a ideia de que a linguagem é o sistema de informação mais fundamental que possuímos, responsável por armazenar e permitir acesso à informação, podemos observar a construção social da linguagem oral e escrita, que são tecnologias tradicionais de informação (Gracioso, Saldanha, 2011). Isso leva Gracioso e Saldanha (2011) a explorarem a base pragmática<sup>2</sup> de Wittgenstein e sua relação circunstancial com a epistemologia da Ciência da Informação, especialmente com a tradição pragmática da epistemologia informacional. Assim como nas narrativas do pragmatismo informacional, os autores não buscam uma definição

---

<sup>2</sup> A pragmática busca entender como a linguagem é usada na comunicação cotidiana, considerando não apenas as palavras em si, mas também o contexto em que são empregadas (Gebauer, 2013)

definitiva de informação, mas sim procuram compreendê-la a partir dos contextos em que o conhecimento é utilizado.

Ludwig Joseph Johann Wittgenstein (1889-1951) foi um filósofo austríaco que teve um impacto significativo na filosofia da linguagem e da mente. Ele é conhecido por suas duas principais obras: *Tractatus Logico-Philosophicus* e "Investigações Filosóficas". No *Tractatus Logico-Philosophicus*, Wittgenstein defende a ideia de que a linguagem tem uma estrutura lógica que reflete a estrutura do mundo. Ele argumenta que a função primordial da linguagem é descrever fatos, e que proposições só fazem sentido se puderem ser verificadas empiricamente. Wittgenstein acreditava que a filosofia deveria lidar com questões lógicas e metafísicas, mas que muitos problemas filosóficos surgem de confusões linguísticas. Nas "Investigações Filosóficas", Wittgenstein abandona muitas das ideias do "Tractatus" e adota uma abordagem mais pragmática da linguagem. Ele argumenta que o significado das palavras não está em sua correspondência com objetos do mundo, mas sim em seu uso dentro de um contexto social. Wittgenstein enfatiza a importância dos jogos de linguagem e das formas de vida na compreensão do significado. Em ambos os trabalhos, Wittgenstein coloca ênfase na importância da linguagem e do uso da linguagem na compreensão e na interação humana. Seu pensamento influenciou o desenvolvimento da filosofia da linguagem, da filosofia da mente e da filosofia da ciência, e suas ideias continuam sendo debatidas e exploradas até os dias de hoje (Gebauer, 2013).

O uso da linguagem pode ser percebido no âmbito da comunicação, seja este natural, artificial ou ambos. Kobashi (2016, p. [1]), ao analisar a Teoria Matemática da Comunicação ou Teoria da informação de Shannon e Weaver (1949) enquanto um modelo, aponta que este "foi amplamente aceito para teorizar a comunicação entre duas máquinas, dois seres humanos ou entre uma máquina e um ser humano". Para tanto, são necessários: fonte de informação, informação, sinal, mensagem, código, transmissor, canal, receptor, destino e utilização de linguagem específica que ocasione economia de tempo e segurança eficazes para contornar possíveis ruídos na transmissão e recepção da mensagem (Kobashi, 2016).

No campo da Ciência da Informação e de suas aproximações com a Biblioteconomia e a Organização e Representação do Conhecimento (ORC),

pode-se citar as Linguagens Documentárias (LD) enquanto sistemas ou instrumentos de representação temática da informação que funcionam como uma ponte entre a Linguagem Natural (LN) do usuário e a linguagem de documentos armazenados ou referenciados em Sistemas de Recuperação da Informação (SRI) (Cintra *et al.*, 2002; Barbosa, 1969; Currás, 2010). As LD podem ser denominadas como Sistemas de Organização do Conhecimento (SOC) – em inglês *Knowledge Organization System* (KOS) –, porque são autênticos instrumentos que viabilizam a representação e recuperação temática da informação, dentre os quais pode-se destacar os sistemas de classificação, as taxonomias, cabeçalhos de assunto, índices, vocabulários controlados e tesouros (Barbosa, 1969; Cintra *et al.*, 2002; Currás, 2005; Santos; Moreira, 2018). Os tesouros e os sistemas de classificação são as LD mais conhecidas e amplamente utilizadas. (Cintra, 2002).

Os SOC do tipo tesouro compreendem um conjunto controlado de termos que representam conceitos pertencentes à determinada terminologia ou domínio do conhecimento (Barros, 2004; Cintra *Et Al.*, 2002; Currás, 2010; Tálamo *et al.*, 1992). Os tesouros são concebidos como as LD mais modernas, que contemplam basicamente três tipos de relações entre os termos: relação hierárquica, relação de equivalência e relação associativa (Cintra, 2002). Deste modo, o tesouro se configura como um vocabulário controlado que, a partir da suposição de modelos ideológicos e socioculturais determinados, tem sua organização hierárquica fundamental funcionando de forma dicionarial. Essa organização “pressupõe uma definição do tipo ‘se > então’, calcada num código de natureza institucional: seus limites dependem dos contratos de natureza social” (Tálamo *et al.*, 1992, p. 198).

Os tesouros, assim como todos os SOC, se constituem em instrumentos imprescindíveis para a organização e representação temática de documentos de um acervo analógico, digital ou híbrido, condicionando a aproximação e recuperação da informação por assunto. No entanto, em comparação a maioria dos SOC, os tesouros são os vocabulários controlados que mais se alinham com o modelo de dados *Simple Knowledge Organization System* (SKOS). Além de serem os únicos SOC que possuem uma norma específica relativamente recente para sua estruturação e mapeamento, a norma ISO 25964 (parte 1:2011 e parte 2:2013). Nesta perspectiva, a presente pesquisa focou nos SOC do tipo tesouros.

A Recuperação da Informação (RI) evolui e se expandiu bastante com o passar das décadas. Inicialmente estática, hoje se apresenta de forma mais interativa ao lidar com diferentes mídias e necessidades informacionais que perpassam, inclusive, o ambiente Web, onde os avanços também são contínuos e especificados em detalhes algorítmicos que permitem que os dados organizados sejam recuperáveis de acordo com os mecanismos de busca (Saracevic, 2009). Assim, elementos como relevância, algoritmos e testes também são cruciais para os processos de RI. A relevância remete ao fato de que as necessidades informacionais dos usuários também devem ser atendidas, ao passo que a recuperação não é sobre qualquer tipo de informação, mas sim sobre informações relevantes, que são de interesse do usuário. Já os algoritmos, que podem ser basicamente definidos como procedimentos lógicos passo a passo, cada vez mais complexos, também contribuem para a construção dos SRI, servindo para estruturação dos mecanismos de busca e automatização de diversas etapas dos fluxos de pesquisa em ambientes digitais. Nesse sentido, toda essa tecnologia algorítmica precisa ser testada e validada em benefício dos usuários (Saracevic, 2009).

A quantidade extraordinária de dados e informações geradas e distribuídas na Web e outros ambientes digitais tem origens e formatos diversos, promovendo, assim, o surgimento do fenômeno denominado *Big Data* e, como dados são gerados através de usos tecnológicos e demandas sociais, este é um fenômeno impulsionador de diferentes necessidades de aprimoramento e adequação das etapas de produção, consumo, compartilhamento e recuperação de informação, com a finalidade de que estas sejam cada vez mais velozes e dinâmicas (Catarino, 2014; Sousa; Martins; Ramalho, 2018b). Assim, constata-se que a identificação de descritores, palavras-chave, *tags*, termos correlatos e sentenças para gerar resultados de busca são necessários para a indexação de páginas Web a partir de *browsers* existentes. Esse tipo de identificação faz parte do arcabouço teórico e prático da Biblioteconomia e Ciência da Informação que lida com indexação manual, automática ou semiautomática em uma variedade de SRI, desenvolvendo formas distintas de representação de documentos, dentre as quais estão os sistemas de classificação, tesauros, “referências, descrição bibliográfica, indexação, resumos, cabeçalhos de assunto, vocabulários controlados, sumários, índices catálogos entre

outras formas que sempre remetem ao documento original” (Sousa; Martins; Ramalho, 2018a, p. 66).

Neste cenário, percebe-se a necessidade de identificar e compreender teoricamente os processos contemporâneos de desenvolvimento de SOC que compõem os processos de RI na Web, em especial de SOC do tipo tesouro. O alcance de tal compreensão contribuirá com o campo da Ciência da Informação a partir do estudo de recomendações para construção de instrumentos de representação temática da informação legíveis por humanos, máquinas e interoperáveis entre si.

Em 2009, o *Word Wide Web Consortium* (W3C) formalizou a recomendação do *Simple Knowledge Organization System* (SKOS) – Sistema de Organização do Conhecimento Simples – como um modelo de dados<sup>3</sup> para a representação de SOC em ambientes digitais, buscando atender às novas demandas informacionais relacionadas ao compartilhamento e interoperabilidade semântica de vocabulários no ambiente Web (Ramalho, 2015; 2017). O modelo de dados SKOS favorece a “evolução das linguagens documentárias e a integração de tesouros em Sistemas de Recuperação de Informação” (Morais; Ramalho; Sousa, 2019, p. 190), uma vez que o W3C considera que dados estruturados e não estruturados na Web devem estar disponíveis e recuperáveis a partir da implementação de dados linkados (*Linked Data*) (Catarino, 2014).

Santos e Moreira (2018) transmitem que o SKOS é uma representação dos SOC enquanto “instrumentos que desempenham, dentre outras, a função de controle de vocabulário”. Para tanto, no modelo SKOS são destacados três elementos principais: os conceitos, as propriedades e as relações. Os conceitos são componentes cruciais na estrutura do SKOS, pois carregam as ideias que expressam os significados de objetos, pessoas, lugares e eventos presentes no mundo real. As propriedades são como os adjetivos que descrevem as características (atributos) dos conceitos e são divididas em duas categorias principais: Propriedades de Etiquetagem e Propriedades de Documentação. As relações expressam a ligação de significado entre os conceitos. São as relações que

---

<sup>3</sup> Pastor-Sánchez (2013a) é um dos autores que utiliza o termo ontologia para se referir ao modelo de dados SKOS, porém, neste trabalho fora adotado o termo modelo de dados, pois é deste modo que a maioria dos textos discutidos na Seção 5 se referem ao SKOS.

atribuem semântica ao SOC representado de acordo com o modelo SKOS e, assim como nos tesouros, podem ser divididas basicamente em três tipos: relações hierárquicas, relações de equivalência e relações associativas (Morais; Ramalho; Sousa, 2019; Ramalho, 2015; Ramalho, 2017).

O modelo SKOS é uma valiosa possibilidade de apresentação formal de vocabulários controlados interoperáveis semanticamente, com estrutura legível por máquina e alternativa de baixo custo para a criação ou conversão de SOC tradicionais em prol da recuperação semântica da informação em ambiente Web. Deste modo, algumas tecnologias semânticas são fundamentais para a estruturação de vocabulários nesse modelo. Os elementos do modelo SKOS são declarados a partir de diretivas triplas *Resource Description Framework* (RDF) e sua estrutura construída com base na linguagem *eXtensible Markup Language* (XML), sendo os conceitos identificados por *Uniform Resource Identifier* (URI) (Catarino, 2014; Miles; Bechhofer, 2009; Ramalho, 2015; Ramalho, 2017; Ramalho; Sousa, 2019).

A principal função do URI é assegurar que os conceitos não sejam reconhecidos apenas como palavras quaisquer, mas que sejam identificados de forma unívoca, com único significado que qualquer indivíduo ou máquina poderá encontrar na Web. Tanto o XML como o RDF são tecnologias semânticas de representação. O XML, enquanto uma linguagem de marcação extensível e versátil, é um formato de representação para qualquer tipo de documento, amplamente utilizado em *softwares* de aplicação Web e recomendado pela W3C desde 1998. O XML possibilita a utilização de *tags* personalizadas, mas não admite erros em seu escopo. Já o RDF é um padrão de representação possível a partir do XML, tendo como objetivo principal promover o intercâmbio de informações entre aplicações sem a perda de significado e, assim, viabilizar a definição de um vocabulário num padrão de representação que permite a interoperabilidade de aplicações (Sousa; Ramalho; Martins, 2018a).

A representação do SKOS em RDF é decorrente da necessidade de ser legível por máquinas e proporcionar a interoperabilidade. As triplas RDF (*subject - predicate - object*) que permitem a codificação SKOS condicionam a representação de elementos dos tesouros para serem lidos e interpretados por sistemas informatizados de maneira interoperável. “Dessa maneira, os relacionamentos são

identificados e tipificados, agregando refinamento semântico nos relacionamentos, o que facilita a leitura destes por computadores e, assim, contribui para automatizações” (Berti Junior *et al.*, 2017, p. 379).

Esse refinamento semântico dos relacionamentos é um dos elementos que agregam qualidade aos tesouros codificados em SKOS. A qualidade dos vocabulários controlados também é implicada pela qualidade dos processos de construção destes, pela eficiência e eficácia do SRI como um todo e pela afinidade da respectiva comunidade usuária com os diferentes tipos de vocabulários controlados. Neste sentido, as normas, padrões e recomendações, bem como as pesquisas de cunho teórico, prático e científico, também são fundamentais para o desenvolvimento de SOC representados em SKOS.

### 1.1 Problematização

Sousa (2019, p. 64) aponta que “com SKOS é possível criar uma série de relações semânticas para estabelecer vínculos entre conceitos diferentes”. Como os tesouros são tipos de SOC, isto implica que, como sistemas de organização suscetíveis de serem desenvolvidos conforme o modelo SKOS, eles carregam a possibilidade de expressão de conceitos representados por termos relevantes e agrupados de forma lógica, onde as categorias conceituais e terminológicas podem ser organizadas de acordo com sua ordem semântica legível por algoritmos. Esta potencialidade dos tesouros representados em SKOS permitiu que a *International Standard Organization* (ISO) “reconhecesse o SKOS, por possuir equivalências terminológicas com os tesouros, quando atualizou a ISO 2788 de 1974 para a ISO 25964 (2011) e (2013)” (Sousa, 2019, p. 64).

A norma ISO 25964 *Thesauri and interoperability with other vocabularies*, Part 1: *Thesauri for information retrieval*, de 2011, e Part 2: *Interoperability with other vocabularies*, de 2013, substituiu as normas anteriores para elaboração de tesouros – como as ISO 2788:1986 e ISO 5964:1985 (Clarke; Zeng, 2012) –, evidencia a importância dos tesouros conceituais em detrimento dos tesouros baseados em termos e fornece recomendações para o desenvolvimento e manutenção de tesouros para recuperar documentos de todos os tipos, independentemente da mídia. A norma também concede orientações para o estabelecimento e manutenção

de mapeamentos entre tesauros e outros tipos de vocabulários, padronizando, assim, aspectos essenciais para a interoperabilidade semântica entre os SOC em sistemas de recuperação da informação (International Standard Organization, 2011; 2013).

Como a referida norma surge diante do fato de que as recomendações SKOS já estavam sendo amplamente utilizadas para construção de tesauros interoperáveis e outros tipos de vocabulários controlados e linguagens documentárias (Bandyopadhyay; Mukhopadhyay, 2015; Martínez-González; Alvite-Díez, 2019a,c), “esta atualização normativa dá ao SKOS uma característica ímpar como modelo padronizado, no qual possui equivalências terminológicas e relacionais com os tesauros” (Sousa, 2019, p. 65). No entanto, o modelo SKOS permite uma integração semântica de vários tipos de SOC no ambiente Web, reconhecendo tecnologias semânticas RDF e XML não apenas para a representação de tesauros e respectivas especificidades (Martínez-González; Alvite-Díez, 2019a,b; Pastor-Sánchez, 2013b), enquanto a norma ISO 25964 partes 1:2011 e 2:2013 volta-se especificamente para a elaboração de tesauros. Sendo assim, esta pesquisa propôs responder ao seguinte questionamento: **Quais as similaridades entre o modelo SKOS e a norma ISO 25964 para a representação de tesauros?**

## 1.2 Justificativas e Hipóteses

Como base para as justificativas e hipóteses da pesquisa realizada, dois fatores foram considerados. O primeiro é que a referida norma surge quando o modelo de dados SKOS já estava sendo amplamente utilizado para construção de tesauros interoperáveis. O segundo é que as recomendações SKOS contemplam todos os tipos de SOC, enquanto as recomendações ISO 25964 visam essencialmente os tesauros. Esses dois fatores indicam possíveis similaridades entre as recomendações SKOS e as recomendações da norma ISO 25964. Assim sendo, mapear e analisar elementos que compõem o modelo SKOS e a norma ISO 25964 contribui no desenvolvimento de instrumentos de representação temática como os tesauros conceituais interoperáveis. Além disso, os padrões e formatos de interoperabilidade são essenciais para o aperfeiçoamento de ambientes semânticos, assim como para a sobrevivência de tesauros já construídos, garantindo a

disseminação, compartilhamento e expansão destes e outros tipos de SOC no ambiente Web.

Atingir a meta de interoperabilidade em redes em contínua expansão demanda esforço substancial e infinitamente escaláveis. Essa questão pode ser analogamente concebida como um quebra-cabeça de padrões e protocolos, em que cada elemento é projetado para se integrar harmoniosamente com as peças vizinhas. Adicionalmente, é necessário considerar tal empreendimento como uma iniciativa comunitária, na qual os desenvolvedores de cada componente do quebra-cabeça devem colaborar mutuamente para assegurar o fluxo contínuo de dados, isento de erros (National Information Standards Organization, 2023). Por conseguinte,

Vizinhos especialmente próximos no quebra-cabeça da Web Semântica são a ISO 25964 e o SKOS. A ISO 25964-1 aconselha essencialmente sobre a seleção e a combinação de conceitos, termos e relacionamentos para criar um bom tesouro. O SKOS aborda o próximo passo, fornecendo recomendações sobre a transferência desses tesouros resultantes (ou outros 'Sistemas de Organização do Conhecimento' simples) para a Web. O ISO 25964-2 recomenda o tipo de mapeamento que pode ser estabelecido entre um SOC e outro; o SKOS apresenta uma forma de expressar esses mapeamentos quando publicados na Web (National Information Standards Organization, 2023, p. [1], tradução livre).

Atentando-se para a expressão “da Web Semântica”, assim como para outras expressões amplamente presentes na literatura dos últimos cinco anos e anterior a este período, tais como “na Web Semântica” e “para a Web Semântica”, torna-se plausível esclarecer que neste trabalho fora considerada, também, a expressão “ambientes semânticos da Web” para se referir à Web Semântica. Isto porque esta extensão da Web é apenas uma parte da rede. “Ambientes semânticos” ou Web Semântica integram tecnologias que se prestam para a configuração em constante desenvolvimento de ambientes semânticos da Web, esta que por sua vez não é totalmente semântica. Desta perspectiva, movimentos e princípios que fazem parte do escopo de ambientes semânticos, como o *Linked Data* e *Linked Open Data* também não são predominantes na Web.

Nesse contexto, a pesquisa se justifica no potencial de contribuição com a produção e compartilhamento de conhecimento científico sobre tesouros interoperáveis em ambientes semânticos da Web no contexto da Organização do

Conhecimento. Justifica-se, também, na importância de evidenciar a relevância de representar tesouros em SKOS.

Os primeiros estímulos para a pesquisa advém de percepções profissionais, científicas, sociais e pessoais no decorrer de atuação da autora enquanto: 1) bibliotecária atuante concomitantemente em organização da informação e serviço de referência da informação (por meio da prestação de serviços de mediação, treinamento e capacitação de usuários da informação); 2) mestranda integrante do Grupo de Pesquisa Organização do Conhecimento e Humanidades Digitais, Linha de Pesquisa “Dados, Informação e Conhecimento no Ambiente Digital”; e 3) ativista social na periferia a partir do trabalho de coordenação democrática do Movimento Negro de Heliópolis e região.

Assim, as justificativas de pesquisa que advém do contexto de atuação enquanto bibliotecária compreendem o fato de que os SOC do tipo tesouro sempre foram domínio dos profissionais da informação, especialmente no âmbito da Biblioteconomia. Porém, a identificação de estudos acerca de normas, padrões e recomendações destes, em conformidade com tecnologias semânticas, é fundamental também para a inserção das bibliotecas nos ambientes da Web Semântica enquanto uma extensão da Web – seja por meio de repositórios digitais ou catálogos OPAC (*Online Public Access Catalog*), por exemplo –, fazendo com que diferentes processos de recuperação da informação disponibilizadas por bibliotecas e outros tipos de unidades de informação sejam viabilizados semântica e dinamicamente no contexto digital dos ambientes semânticos da Web. Nesse sentido, também, a elaboração de manuais e guias de procedimentos para construção de SOC do tipo tesouro conforme o modelo SKOS pode ser acorçoada, pois junto aos manuais e normas já existentes para tal é fundamental considerar as recomendações estabelecidas para a construção e uso de SOC legíveis por máquinas e humanos.

Quanto ao segundo contexto de atuação, enquanto mestranda, a justificativa para identificar cientificamente as similaridades entre o modelo SKOS e a norma ISO 25964 para a representação de tesouros se constitui no fato de que esta identificação é tarefa primordial para apoiar pesquisas e inovações para a formulação conceitual, construção, conversão e uso de tesouros para emprego em

diferentes contextos de recuperação da informação que agreguem ainda mais ao campo da informação.

Neste sentido, algumas propostas e contextos de aplicação e desenvolvimento de tesouros codificados em SKOS foram identificadas a partir de levantamento bibliográfico prévio à execução da presente pesquisa, como por exemplo: recuperação da informação por meio de aplicações de Web Semântica atreladas à Inteligência Artificial (IA) via Processamento de Linguagem Natural (PLN) em *chatbots* (Coneglian, 2020) e; processos de desenvolvimento de *softwares* com vistas a padronizar vocabulários de desenvolvimento do produto digital (Zacarias, 2020), apoiando o trabalho de desenvolvedores de *software* e *Product Owners* (PO). Assim, estas e outras possibilidades de aplicação de tesouros representados em SKOS convergem com o primeiro e segundo contexto de justificativas para esta pesquisa – atuação profissional (bibliotecária) e científica (mestranda) –, pois também sinalizam a possibilidade de diversos ambientes em que aplicações e investigações científicas acerca dos tesouros representados em SKOS, em conformidade com a normalização e recomendações vigentes (em benefício de enriquecimento semântico em recuperação da informação, a partir da representação temática legível por máquina e humanos) possam ser ponderadas: ambientes digitais de arquivos, museus, repositórios digitais, sistemas de gerenciamento eletrônico de documentos, catálogos OPAC, Ambientes Virtuais de Aprendizagem (AVA) e bibliotecas digitais/virtuais temáticas. Ademais, os resultados da pesquisa apresentados neste trabalho também podem contribuir para a revitalização da atuação de profissionais da informação em diversos nichos de negócios a serem favorecidos pela oferta de produtos e serviços de informação potencializados pela representação temática via tesouros representados em SKOS e devidamente normalizados.

No que tange às justificativas que advém do terceiro contexto de atuação, enquanto ativista social, o que se percebe é que os usuários informacionais reais e potenciais, de diferentes unidades de informação e respectivos produtos e serviços disponíveis na Web, intencionam processos de pesquisa cada vez mais dinâmicos e significativos para suprir necessidades informacionais específicas, demonstrando a importância de que as relações semânticas entre os assuntos dos recursos de informação estejam cada vez mais refinadas em benefício da agilidade no alcance

de resultados de busca relevantes em SRI. Neste sentido, os tesouros enquanto instrumentos terminológicos de indexação e apoio à pesquisa em SRI podem ser integrados à RI de forma interoperável e permissível a “descoberta acidental de conteúdos” (Castro; Jesus, 2018, p. 151), de modo a nortear usuários informacionais que detenham ou não aprendizado prévio à realização das tarefas de busca e a termos representativos de assuntos e conceitos de determinada área do conhecimento em ambientes digitais e semânticos da *Web* (Baeza-Yates; Ribeiro-Neto, 2013).

Ainda sobre as justificativas sociais para realização desta pesquisa, percebe-se que os tesouros representados em SKOS enquanto instrumentos de representação e recuperação de informação de terminologias específicas, precisam contemplar, também, domínios do conhecimento contra-hegemônicos, como é o caso dos saberes advindos de povos indígenas e afro-brasileiros. Haja vista que estes grupos possuem necessidades informacionais específicas, também se beneficiariam da integração de tesouros semânticos e interoperáveis em SRI. Nesta perspectiva, espera-se que a presente pesquisa fundamente abordagens e aplicações futuras de tesouros representados em SKOS voltados a domínios contra-hegemônicos do conhecimento.

Pode-se afirmar que as justificativas que advém do contexto de atuação social da autora se congregam não apenas como justificativas sociais, mas também pessoais, pois a autora desta pesquisa se reconhece enquanto cidadã afro-brasileira que cotidianamente se depara com diversos obstáculos de orçamento para desenvolver pesquisas como esta, assim como para desenvolver instrumentos de recuperação da informação mais adequados às minorias sociais com o qual convive e atua. Por isso, chama atenção a constatação de que “o SKOS surge como uma ferramenta de baixo custo que é capaz de compreender a semântica nativa de uma variedade de KOS através de um conjunto de propriedades e, assim, realizar a migração destes materiais para a *Web*” (BAKER *et al.*, 2013, p.37 *apud* Moraes; Ramalho; Sousa, 2019, p. 191), o que também torna este modelo atrativo para o condicionamento da recuperação semântica da informação, inclusive, de fontes de informação que salvaguardam a história das lutas e conquistas empregadas por movimentos sociais.

As problemáticas que possivelmente decorreriam da não resolução do problema desta pesquisa se constituem na inviabilidade de análises científicas e técnicas fundamentais para a detecção das relações entre a norma ISO 25964, parte 1:2011 e 2:2013, e as recomendações do modelo de dados SKOS que, por conseguinte embasam estudos e desenvolvimentos de abordagens e aplicações estritamente relacionadas ao SOC do tipo tesouro representados em SKOS devidamente normalizados. Porque mesmo que os tesouros sejam compreendidos como instrumentos de representação temática fundamentais para a representação, organização e recuperação da informação, a não identificação e análise metodológica dos estudos e práticas entorno das recomendações e normalização vigentes destes instrumentos semânticos, adaptados para o contexto digital da Web, compromete o alcance de resultados de pesquisa que embasam os estímulos e possibilidades de novos estudos e desenvolvimento amplo de novos tesouros representados em SKOS e normalizados, conforme a ISO 25964. Ademais, o estudo acerca das similaridades entre o modelo SKOS e norma ISO atual para representações de tesouros poderá apoiar novas abordagens e integrações dos tesouros aos processos entorno de: preservação digital; representação temática da informação no contexto do *Linked Open Data* (LOD); engenharia de tesouros interoperáveis; refinamento semântico; recuperação da informação digital de domínio contra-hegemônico ou não; inteligência artificial; desenvolvimento de produto digital; indexação em arquivos, museus e bibliotecas e; preservação de memória institucional.

### 1.3 Objetivos

O objetivo geral da pesquisa é apresentar similaridades entre o modelo SKOS e a norma ISO 25964<sup>4</sup> para a representação de tesouros. Para tanto, os objetivos específicos são:

- a) Analisar a estrutura do modelo de dados SKOS, recomendações e norma relacionada, atentando-se para especificidades e similaridades com a estrutura de tesouros clássicos;

---

<sup>4</sup> Acesso à norma impressa ISO 25964 completa se deu por meio da Biblioteca da Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP).

- b) Apontar aspectos normativos em conformidade com a norma ISO 25964 para a representação de tesouros conceituais e interoperáveis;
- c) Verificar abordagens de aprimoramento, extensões e usos no âmbito do cenário geral de aplicações conjuntas do modelo SKOS e da norma ISO 25964.

#### **1.4 Abordagem metodológica**

Conforme conceitos de metodologia do trabalho científico esclarecidos por Severino (2009), Gerhardt e Silveira (2009), a presente pesquisa possui abordagem qualitativa de natureza aplicada, pois visa produzir conhecimentos que embasam a aplicação prática, dirigidos à solução do problema de pesquisa anunciado anteriormente. Está caracterizada como exploratória e descritiva, pois para o alcance de seus objetivos geral e específicos pretende proporcionar maior familiaridade com o problema e resolução deste a partir da análise, identificação e descrição dos aspectos do objeto estudado, de modo a contribuir com a identificação do estado da arte nacional e internacional das abordagens e aplicações de SOC do tipo tesouro representados em SKOS. Quanto aos procedimentos utilizados, esta pesquisa é bibliográfica, porque é realizada a partir do levantamento de referências formalmente elaboradas, publicadas e analisadas em artigos científicos, livros, relatórios científicos, documentos, normas e páginas Web institucionais ou oficiais. Contudo, o centro das fontes analisadas foi dado aos artigos científicos indexados em bases de dados nacionais e internacionais.

A análise das fontes ocorreu por meio do método de análise de conteúdo categorial. Para tanto, toda a conceituação e procedimentos metodológicos desta pesquisa estão enquadrados no Modelo Metodológico Quadripolar elaborado e proposto pelos cientistas belgas Paul de Bruyne, Jacques Herman e Marc de Schoutheete. Este enquadramento permitiu que a investigação fosse conduzida dinamicamente por meio dos quatro polos metodológicos que constituem o Modelo Metodológico Quadripolar: polo epistemológico, polo teórico, polo morfológico e polo técnico (Bruyne; Herman; Schoutheete, 1977).

## 1.5 Estruturação da pesquisa

A estrutura da presente dissertação está definida conforme os objetivos da pesquisa. Então, após esta primeira Seção introdutória, a *Seção 2 Procedimentos metodológicos: Modelo Metodológico Quadripolar* é composta pela apresentação dos procedimentos metodológicos enquadrados no Modelo Metodológico Quadripolar, o que inclui a definição e fundamentação das escolhas e instrumentos metodológicos que amparam os objetivos propostos.

A *Seção 3 Encadeamentos entre Ciência da Informação e Tesouros* apresenta uma contextualização dos conceitos fundamentais para o entendimento da estrutura dos tesouros clássicos à luz da Ciência da Informação, compreendendo: Linguagem, Linguagens Documentárias, Informação, Representação Temática da Informação, Recuperação da Informação e Tesouro. Estes conceitos fundamentam a abstração da estrutura dos tesouros representados em SKOS e discuti-los permite a apresentação do universo de pesquisa, bem como a descrição do estado da arte do objeto estudado – modelo de dados SKOS para representação de tesouros, em consonância com a norma ISO 25964 *Thesauri and interoperability with other vocabularies, Part 1: Thesauri for information retrieval* (2011) e *Part 2: Interoperability with other vocabularies* (2013).

A *Seção 4 Simple Knowledge Organization System e norma ISO 25964* discorre sobre a estrutura do modelo de dados SKOS a partir do contexto em que foi desenvolvido, recomendações e norma vigente para representação de tesouros, atentando-se para especificidades e similaridades entre a estrutura SKOS e a estrutura de tesouros clássicos.

Enquanto resultados apresentados e discutidos, as similaridades entre o modelo SKOS e a norma ISO 25964 são apresentadas na *Seção 5 Análise comparativa entre a norma ISO 25964 e o modelo SKOS*, em subseções que favorecem os pontos de similaridades levantados a partir do estudo da norma, do modelo de dados e da bibliografia levantada.

As considerações finais, presentes na *Seção 6*, são estabelecidas no destaque das principais conclusões sobre os resultados alcançados, de modo a

trazer apontamentos sobre as fragilidades e potencialidades deste trabalho e sugestões de pesquisas futuras.

## **2 PROCEDIMENTOS METODOLÓGICOS: MODELO METODOLÓGICO QUADRIPOLAR**

O percurso da pesquisa científica relatada nesta dissertação é guiado pela aplicação do Modelo Metodológico Quadripolar. Deste modo, esta seção abarca as nuances epistemológicas e práticas metodológicas percorridas. Na subseção 2.1 é apresentado o Modelo Metodológico Quadripolar para o enquadramento metodológico, sendo a subseção 2.2 dedicada à exposição dos procedimentos metodológicos técnicos.

### **2.1 Enquadramento metodológico: Modelo Metodológico Quadripolar**

A pesquisa apresentada nesta dissertação se respalda no método quadripolar, que condiciona estratégia científica flexível e dinâmica para aplicação dos procedimentos metodológicos realizados.

No âmbito da estratégia metodológica convencional para pesquisas em Ciências Sociais tem-se na maioria dos casos apenas dois polos: o referencial teórico e a pesquisa de campo ou bibliográfica e documental. No primeiro polo estão integradas as categorias de pesquisa vistas sob o enfoque teórico que às vezes observa aspectos epistemológicos. O segundo polo convencional é o que estrutura a pesquisa, dá ênfase à metodologia científica, o levantamento, análise, resultados e discussão. Já a metodologia quadripolar utiliza quatro polos.

O Modelo Metodológico Quadripolar é uma proposta de abordagem qualitativa para pesquisas científicas de diferentes áreas das Ciências Sociais. Este modelo foi elaborado e proposto pelos cientistas belgas Paul de Bruyne, Jacques Herman e Marc de Schoutheete que o consagraram a partir da publicação do livro *Dynamique de la recherche en sciences sociales de pôles de la pratique méthodologique*, no ano de 1974. A edição em português brasileiro, intitulada *Dinâmica da pesquisa em Ciências Sociais: os pólos da prática metodológica*, foi publicada em 1977.

Machado *et al.* (2016) exploram o Modelo Metodológico Quadripolar em destaque à sua importância para as pesquisas de abordagem qualitativa, haja vista o desafio de investigar fenômenos sociais sem cair na objetivação do que é

subjetivo, sem perder a garantia de manutenção da cientificidade de diversas formas de investigação e conjunto de técnicas comuns à pesquisa qualitativa, que tem resultados descritivos como produto. Embora o Modelo Metodológico Quadripolar não seja muito divulgado, sua importância se destaca diante de sua consistência para atribuição de cientificidade aos processos de pesquisas voltadas à sociedade, indivíduos e seus respectivos fenômenos através de quatro polos metodológicos que podem ser utilizados tanto na investigação quantitativa quanto qualitativa, são eles: polo epistemológico, polo teórico, polo morfológico e polo técnico (Machado *et al.*, 2016; Silva, 2014;).

O polo epistemológico é o contexto em que o pesquisador irá levantar os fundamentos do fenômeno ou das categorias teóricas ou das categorias de pesquisa que ele irá investigar e explorar. Este polo é responsável pelas regras de produção e de explicação dos fatos, da compreensão e da realidade das teorias que regem as regras de transformação do objeto de estudo científico propriamente dito. Neste polo estão as problemáticas da pesquisa concernentes ao paradigma considerado para a pesquisa. Isto vem de uma visão de mundo, uma visão de humanidade que culminaram numa visão de ciência, em que a pesquisa será identificada com uma ou mais abordagens, pois a epistemologia contemporânea traz a possibilidade de correlação. No polo epistemológico o problema é tratado e fundamentado. Neste polo é dada a garantia de objetivação e vigilância crítica da pesquisa (Bruyne; Herman; Schoutheete, 1977; Machado *et al.*, 2016; Silva, 2014).

O polo teórico é o palco dos conceitos, onde será explicitada a parceria de teóricos, matrizes teóricas e escolas teóricas que referenciam a pesquisa. O polo teórico é o guia de elaboração de hipóteses e da construção dos conceitos. Neste polo são propostas regras de interpretação dos fatos, de especificações, de definição das soluções provisórias dadas à problemática. No polo teórico está o quadro de referência de teóricos que estarão fundamentados pela base epistemológica. Neste polo está o campo abstrato, no qual procurar-se-á fundamentar e conceituar as teorias da pesquisa, se utilizando de teóricos e correntes teóricas sobre as categorias pesquisadas (Bruyne; Herman; Schoutheete, 1977; Silva, 2014; Machado *et al.*, 2016). Este polo “é o lugar de elaboração das linguagens científicas, determina o movimento de conceitualização” (Machado *et al.*, 2016, p. 278).

O polo morfológico consiste na integração da teoria ao real, da teoria à prática. Neste polo estão os modelos, as regras de estruturação, formação e ordenação dos elementos do objeto científico. O polo morfológico é a instância das regras de estruturação, apresentação dos dados e de formação do objeto científico, de utilização de modelos, de tipologias, de simulacros, de sistemas e quadros de análise. Neste polo são identificados modelos teóricos e operacionais (pode-se dizer que o Modelo Metodológico Quadripolar é um desses modelos teóricos e operacionais). Modelos teóricos são modelos mais conceituais e modelos operacionais são modelos mais aplicados, mais práticos. Aqui é configurada e esquematizada toda a pesquisa a partir de autores que já construíram modelos ou por meio da criação de novos modelos, novas morfologias (Bruyne; Herman; Schoutheete, 1977; Machado *et al.*, 2016; Silva, 2014).

Por sua vez, o polo técnico é o polo de aplicação dos métodos de investigação, propiciando a discussão dos resultados que estão sendo gerados pela pesquisa. Neste polo está a coleta de dados, bem como o esforço por confrontá-los com a teoria que os suscitou. No polo técnico estão os modos de investigação: pesquisa de campo, estudo de caso, estudos comparativos, experimentação e/ou simulação. Neste polo o/a pesquisador/a se volta para o cenário real da pesquisa, pois se o polo morfológico traz modelos e simulacros, o polo técnico abarca o contexto concreto da investigação científica, onde o/a cientista poderá ir a campo realizar a pesquisa. Então, neste polo, o/a pesquisador/a irá investigar, utilizar técnicas e modos de investigação, fará comprovações. Aqui o problema inicial será traduzido num novo problema que poderá ser colocado de outra maneira, numa outra pesquisa ou na continuidade da pesquisa vigente (Bruyne; Herman; Schoutheete, 1977; Machado *et al.*, 2016; Silva, 2014).

No ano de 2002, por meio da publicação do livro *Das “Ciências” documentais à ciência da informação: ensaio epistemológico para um novo modelo curricular*, Armando Malheiro da Silva e Fernanda Ribeiro apontaram o Modelo Metodológico Quadripolar como um método adequado para as pesquisas em Ciência da Informação, que é uma Ciência Social Aplicada cujos processos de pesquisa podem ser facilitados pelo uso deste modelo. Silva e Ribeiro (2002) apresentam um debate entre as ideias de documento e de informação social – “formulada assim para significar que se baseia em signos e símbolos codificados e decodificados por

seres humanos em grupo ou sociedade” (Silva, 2014, p.30). Nesse sentido, a informação pode ser vista enquanto um fenômeno info-comunicacional e objeto de estudo científico que está presente não apenas em diferentes unidades de informação e dispositivos de armazenamento informacional, tais como Bibliotecas, Museus, Arquivos e discos rígidos de computadores, mas também em diferentes disciplinas práticas e profissionais, como a Arquivística, Biblioteconomia, Museologia e Documentação, fazendo com que a informação precise ser conhecida e trabalhada de maneira transdisciplinar, enquanto possibilidade epistêmica que faz da informação objeto científico comum e único, gerador de uma perspectiva que lança um novo campo científico resultante dessa transdisciplinaridade, designado Ciência da Informação (Silva, 2014; Silva; Ribeiro, 2002).

A transdisciplinaridade da Ciência da Informação, então, levou à necessidade de especificação interna dos quatro polos metodológicos da prática científica propostos por Bruyne, Herman e Schoutheete (1977). Assim, no âmbito desta especificação, o polo epistemológico é o que fundamenta os demais polos, pois denota a “instância superior imbricada no aparato teórico e institucional (a comunidade científica dos especialistas em informação, as suas escolas, institutos, locais de trabalho, e ainda os seus referentes políticos, ideológicos e culturais)” (Ribeiro; Silva, 2002, p. 87). No polo epistemológico pode ser apresentado o percurso histórico-evolutivo da Ciência da Informação, suas características enquanto campo científico e o cenário paradigmático em que se configura o objeto de estudo (Vechiato, 2013).

No polo teórico se situa o referencial teórico cuja formação indicará as teorias e respectivos autores que dialogam com o/a pesquisador/a e sua maneira de conhecer o objeto científico e “respectiva postulação de leis, formulação de conceitos operatórios, hipóteses e teorias (plano de descoberta) e subsequente verificação ou refutação do ‘contexto teórico’ elaborado (plano de prova)” (Silva; Ribeiro, 2002, p. 87). Neste polo podem ser abordados os conceitos e respectivas nomenclaturas pertinentes aos elementos que compõem o objeto científico e o contexto histórico e contemporâneo no qual se insere, em conformidade com o *status* e atributos dos conceitos no campo da Ciência da Informação (Vechiato, 2013).

No polo morfológico o objeto científico toma forma de acordo com toda a pesquisa realizada. Neste polo “se assume por inteiro a análise/avaliação e se parte não apenas para a configuração do objecto científico, mas também para a exposição de todo o processo que permitiu a sua construção” (Silva; Ribeiro, 2002, p. 90). Neste polo é exercida a função de comunicação da pesquisa, pois são apresentados e organizados os dados em prol da descrição e/ou modelagem do objeto científico (Bruyne; Herman; Schoutheete, 1977; Silva, 2014). No polo morfológico também é discutido o modelo do objeto científico proposto pelo/a pesquisador/a (Vechiato, 2013).

O polo técnico é onde o/a pesquisador/a se mune de instrumentos e técnicas que viabilizarão o alcance dos objetivos da pesquisa. Neste sentido, os procedimentos técnicos presentes no domínio da Arquivística e da Biblioteconomia acumularam-se de tal modo que impõem revisão do alinhamento das técnicas de representação formal da documentação, armazenamento, transferência, recuperação e difusão da informação, porque estas técnicas tendem a estar avulsas e dispersas e no polo técnico há a necessidade de verificação ou refutação do contexto histórico dos métodos vigentes (Silva, 2014; Silva; Ribeiro, 2002). Enquanto no polo epistemológico cabe também a caracterização da pesquisa com relação a sua abordagem (quantitativa, qualitativa ou mista) e natureza (básica ou aplicada), o polo técnico trata da realização da pesquisa em si, em sua prática que a configura conforme seus objetivos e procedimentos.

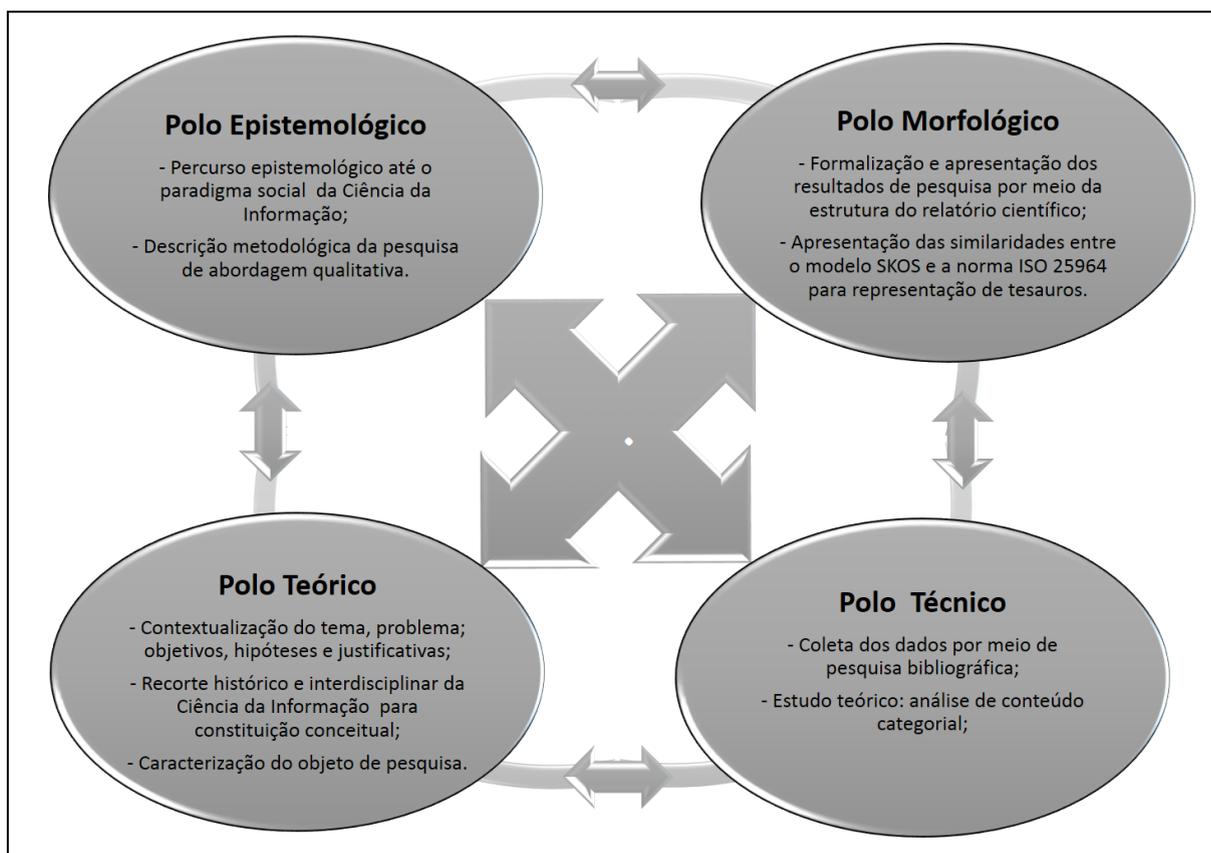
Durante a pesquisa, os quatro polos do Modelo Metodológico Quadripolar interagem entre si, retroalimentando-se de maneira alternada e não-linear, visto que

A prática científica não é redutível a uma sequência de operações, de procedimentos necessários e imutáveis, de protocolos codificados. Tal concepção, que converte a metodologia numa tecnologia, repousa sobre a visão rigorista e “burocrática” do design, fixado no início da pesquisa [...], “mania de projeto”. Ao contrário, parece que a complexidade das problemáticas em ciências sociais exige interpenetrações e voltas constantes entre os pólos epistemológico, teórico, morfológico e técnico da pesquisa”. (Bruyne, Herman, Shoutheete, 1977, p. 30).

A Figura 1 representa a forma como os quatro polos do método quadripolar se configuram e estão ajustados na presente dissertação. Ilustra, portanto, a

interação mútua entre os quatro polos que se retroalimentam durante todo o processo de pesquisa, a saber:

**Figura 1** - Método quadripolar de pesquisa: interação entre os polos.



Fonte: Elaborado pela autora.

O **polo epistemológico** é constituído pelo percurso epistemológico até o paradigma social da Ciência da Informação e descrição metodológica da pesquisa como sendo de abordagem qualitativa de natureza aplicada, pois visa produzir conhecimentos que embasam a aplicação prática, dirigidos à solução do problema de pesquisa previamente anunciado (Gerhardt; Silveira, 2009; Severino; 2009) está caracterizada como exploratória e descritiva, ao passo que para alcance de seus objetivos geral e específicos pretende proporcionar maior familiaridade com o problema e resolução deste a partir da análise, identificação e descrição dos aspectos do objeto estudado (Gerhardt; Silveira, 2009; Severino; 2009), de modo a contribuir com a apresentação das similaridades entre o modelo SKOS e a norma ISO 25964 para representação de tesouros. A forma como os dados foram

analisados (identificada no polo técnico) está atrelada ao método científico hipotético-dedutivo.

Para fundamentar a contextualização do tema, elaboração do problema de pesquisa, objetivos, hipóteses e justificativas, o **polo teórico** se refere ao recorte histórico, conceitual e interdisciplinar da Ciência da Informação em prol da apresentação dos conceitos que caracterizam o objeto de pesquisa e norteiam a investigação científica.

No **polo técnico** estão os procedimentos utilizados (descritos tecnicamente na subseção 4.2) para coleta e análise dos dados. Neste polo, a pesquisa é bibliográfica, porque a coleta dos dados é realizada a partir do levantamento de referências formalmente elaboradas, publicadas e analisadas em artigos científicos, livros, relatórios científicos, documentos, normas e páginas Web institucionais ou oficiais. Contudo, o centro das fontes analisadas foi dado aos artigos científicos indexados em bases de dados nacionais e internacionais, considerando-se que “hoje, predomina entendimento de que artigos científicos constituem o foco primeiro dos pesquisadores, porque é neles que se pode encontrar conhecimento científico atualizado, de ponta” (Marconi; Lakatos, 2022, p. 33). A análise das publicações ocorreu por meio de estudo teórico embasado no método de análise de conteúdo categorial. “A análise de conteúdo vale-se da inferência para extrair as questões relevantes contidas no conteúdo das mensagens” (Valentim, 2005, p. 124). A utilização desta forma de análise está atrelada ao método científico hipotético-dedutivo proposto por Karl Popper, no qual as hipóteses levantadas são verificadas conforme o processo de investigação científica realizado (Marconi; Lakatos, 2022, p. 94).

No **polo morfológico** está a formalização e apresentação dos resultados de pesquisa por meio da estrutura e formatação do relatório científico e das categorias de análise das similaridades entre o modelo SKOS e norma ISO 25964. A própria utilização do Modelo Metodológico Quadripolar se constitui num elemento que integra o polo morfológico, pois este modelo contribuiu para a estruturação (forma/morfologia) da pesquisa.

O Quadro 1 relaciona as seções e respectivos polos que dão forma a esta dissertação:

**Quadro 1 - Estrutura do relatório de pesquisa.**

<b>Seção</b>	<b>Descrição</b>	<b>Polo</b>
1 Introdução	Contextualização do tema, problema, objetivos, hipóteses, justificativas e da estrutura da pesquisa.	Teórico
2 Procedimentos metodológicos: Modelo metodológico quadripolar	Descrição metodológica da pesquisa de abordagem qualitativa;  Coleta dos dados por meio de pesquisa bibliográfica.	Epistemológico, Técnico e Morfológico
3 Encadeamento entre a Ciência da Informação e os tesouros	Percurso epistemológico até o paradigma social da Ciência da Informação;  Recorte histórico e interdisciplinar da Ciência da Informação para constituição conceitual.	Teórico e Epistemológico
4 Simple Knowledge Organization System e norma ISO 25964-1:2011 e ISO 25964-2:2013	Caracterização do objeto de pesquisa.	Teórico
5 Análise comparativa entre o modelo SKOS e a norma ISO 25964	Formalização e apresentação dos resultados de pesquisa por meio da estrutura do relatório científico e das categorias de análise das similaridades entre o modelo SKOS e norma ISO 25964.  Estudo teórico: análise de conteúdo categorial.	Técnico, Morfológico
6 Considerações finais	Resgate reflexivo de elementos dos polos Epistemológico, Teórico, Técnico e Morfológico.	Epistemológico, Teórico, Técnico e Morfológico

Fonte: Elaborado pela autora.

Percebe-se que a estrutura da pesquisa possui seções que se enquadram em mais de um polo do Modelo Metodológico Quadripolar. Este enquadramento

simultâneo é um reflexo da interação dinâmica e não sequencial entre os polos e seus elementos constituintes.

Deste modo, toda a conceituação e procedimentos metodológicos desta pesquisa estão enquadrados no Modelo Metodológico Quadripolar. Este enquadramento permitiu que a investigação fosse conduzida dinamicamente por meio dos quatro polos metodológicos que constituem o Modelo Metodológico Quadripolar: polo epistemológico, polo teórico, polo morfológico e polo técnico (Bruyne; Herman; Schoutheete, 1977).

## 2.2 Descrição dos procedimentos técnicos

Os procedimentos técnicos da metodologia se constituem em 5 etapas, onde os dados coletados constituem o *corpus* de análise composto por registros recuperados a partir de estratégia de busca direcionadas aos campos de título, resumo e palavras-chave. Essas estratégias foram configuradas de acordo com os recursos de busca e filtros disponíveis nas seguintes bases de dados, repositórios digitais ou *sites*: Base de Dados do Encontro Nacional de Pesquisa em Ciência da Informação (BENANCIB); *site* brasileiro com publicações da *International Society for Knowledge Organization* (ISKO Brasil); Web of Science; Scopus; *Library, Information Science & Technology Abstracts with Full Text* (LISA); *Academic Search Premier* (ASP); *Information Science & Technology Abstracts* (ISTA); *Scientific Electronic Library Online* (SciELO.org); Science Direct; *Gale Academic OnFile*; Biblioteca Digital Brasileira de Teses e Dissertações do Instituto Brasileiro de Informação em Ciência e Tecnologia (BDTD/IBICT); e Base Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI). As buscas foram realizadas entre maio de 2021 e maio de 2022 e totalizaram 626 resultados.

A primeira etapa desenvolvida foi formular as estratégias de busca respeitando-se as especificidades de cada base de dados, na tentativa de potencializar o alcance de resultados. Os dois termos pesquisados foram: “SKOS” e “*Simple Knowledge Organization System*”, utilizando o operador booleano OR e considerando os filtros para publicações entre 2012 e 2022, em idioma português, inglês ou espanhol.

O Quadro 2 sintetiza as estratégias de busca e filtros de acordo com a quantidade de publicações recuperadas em cada base de dados, repositório ou *site*:

**Quadro 2** - Estratégias de busca de acordo com bases de dados e resultados.

Base de dados/repositório ou site	Formulação de estratégia com filtro de busca	Quantidade de documentos recuperados
<b>Base de Dados do Encontro Nacional de Pesquisa em Ciência da Informação (BENANCIB)</b>	SKOS (texto completo), pois era a única opção disponível	21 resultados
<b>International Society for Knowledge Organization (ISKO Brasil)</b>	SKOS OR "Simple Knowledge Organization System" via "ctrl+F" em cada um dos pdfs das publicações disponíveis no site	5 resultados
<b>Web of Science</b>	SKOS (Tópico) or "simple knowledge organization system" (Tópico) and Information Science & Library Science (Categorias da Web of Science) and 2022 or 2021 or 2020 or 2019 or 2018 or 2017 or 2016 or 2015 or 2014 or 2013 or 2012 (Anos da publicação) and English or Spanish or Portuguese	188 resultados
<b>Scopus</b>	( TITLE-ABS-KEY ( skos ) OR TITLE-ABS-KEY ( *simple AND knowledge AND organization AND system* ) AND LANGUAGE ( english ) AND LANGUAGE ( spanish ) AND LANGUAGE ( portuguese ) ) AND PUBYEAR > 2011 AND PUBYEAR > 2011	0 resultados
<b>LISA</b>	TI SKOS OR SU SKOS OR AB SKOS OR TI "simple knowledge organization system" OR SU "simple knowledge organization system" OR AB "simple knowledge organization system"  Limitadores - Data de publicação: 20120101-20221231; Idioma: English, Portuguese, Spanish Expansores - Aplicar assuntos equivalentes Modos de busca - Booleano/Frase	81 resultados
<b>ASP</b>	TI SKOS OR SU SKOS OR AB SKOS OR TI "simple knowledge organization system" OR SU "simple knowledge organization system" OR AB "simple knowledge organization system"	148 resultados

	<p>Limitadores - Data de publicação: 20120101-20220531; Idioma: English, Portuguese, Spanish</p> <p>Expansores - Aplicar assuntos equivalentes</p> <p>Modos de busca - Booleano/Frase</p>	
<b>ISTA</b>	<p>TI SKOS OR SU SKOS OR AB SKOS OR TI "simple knowledge organization system" OR SU "simple knowledge organization system" OR AB "simple knowledge organization system"</p> <p>Limitadores - Data de publicação: 20120101-20220531; Idioma: English, Spanish</p> <p>Expansores - Aplicar assuntos equivalentes</p> <p>Modos de busca - Booleano/Frase</p>	43 resultados
<b>SciELO.org</b>	<p>(ti:(SKOS)) OR (ab:(SKOS)) OR (kw: SKOS) OR (ti:("simple knowledge organizations system")) OR (ab:("simple knowledge organizations system")) OR (kw: "simple knowledge organizations system")</p>	4 resultados
<b>Science Direct</b>	<p>Year: 2012-2022</p> <p>Title, abstract, keywords: SKOS OR "simple knowledge organization system"</p> <p>Limitadores - Data de publicação: 2012:2022; Idioma: English, Portuguese, Spanish</p>	51 resultados
<b>Gale Academic OnFile</b>	<p>Termos da pesquisa: Título do documento: SKOSORResumo: SKOSORPalavra-chave: SKOSORTítulo do documento: "simple knowledge organizations system"ORResumo: "simple knowledge organizations system"ORPalavra-chave: "simple knowledge organizations system"</p> <p>Filtros aplicados: Texto completo Revisado por especialistas Data: Jan 01, 2012 - May 30, 2022</p>	23 resultados
<b>BDTD - IBICT</b>	<p>Termos da pesquisa: Busca: (Título:SKOS OU Assunto:SKOS OU Resumo Português:SKOS OU Resumo inglês:SKOS OU Título:"simple knowledge organization system" OU Assunto:"simple knowledge organization system" OU Resumo Português:"simple knowledge organization system" OU Resumo inglês:"simple knowledge organization system")</p>	4 + 1 resultados

	Filtros aplicados: Texto completo Revisado por especialistas Data: Jan 01, 2012 - May 30, 2022 + Bucas Alternativas: (...) "simple knowledge organization systems" (...)	
<b>BRAPCI</b>	Termos da pesquisa (individualmente, em campo título, palavras-chave e Resumo, devido a problemas de configuração da estratégia de busca):  Título:"simple knowledge organization system" or SKOS Assunto:"simple knowledge organization system" or SKOS Resumo: "simple knowledge organization system" or SKOS  Filtros aplicados: Texto completo Revisado por especialistas Data: Jan 01, 2012 - May 30, 2022	58 resultados
<b>Total:</b>		627 resultados

Fonte: Elaborado pela autora.

Na segunda etapa foi elaborada uma planilha no *software* Excel com a finalidade de transportar os dados coletados de acordo com os metadados necessários para execução da análise de conteúdo categorial. Cada coluna da planilha representa o tipo metadado referente para cada publicação (dado) recuperada e disposta em cada linha da planilha. As colunas são: Título; 1º autor; Área de aplicação; Tipo de trabalho; Ano de publicação; Idioma; Paginação; Revista/Publicação/Anais/Evento; Grupo de trabalho; Base de Dados; Estratégia de busca; *Abstract*; Palavras-chave; Tipo de leitura (Não, Integral, Parcial, *Abstract*); Assuntos relacionados; DOI (*Digital Object Identifier*); URL (*Uniform Resource Locator*); Referência ABNT (Associação Brasileira de Normas Técnicas). Além das colunas (metadados) título, *abstract* e palavras-chave, as colunas área de aplicação e assuntos relacionados também permitiram identificar as propostas das publicações, de modo a contextualizar a posterior leitura do documento recuperado.

Da terceira à quarta etapa tem-se o início da realização da análise de conteúdo categorial propriamente dita. Após leitura e análise dos títulos, resumos e palavras-chave dos 627 documentos recuperados, foram identificadas e

selecionadas as publicações que tratavam da relação entre o modelo de dados SKOS e a norma ISO 25964 (partes 1 e 2), resultando num *corpus* inicial de análise composto por 13 artigos científicos. Como parte dos resultados, foram identificadas abordagens e aplicações que envolvem a relação entre a norma ISO 25964 e o modelo de dados SKOS, revelando a necessidade de realização de mais um recorte sobre os 627 resultados iniciais. Este segundo processo de recorte buscou identificar e selecionar para leitura e revisão as publicações que tratavam das principais extensões SKOS, pois estas se constituem na tentativa de ampliar o modelo SKOS de modo a atender as especificidades da norma ISO 25964 para representação de tesouros (Barbosa; Viera, 2022). Deste modo, buscou-se, nos títulos, palavras-chave e resumos, menções às extensões SKOS-XL, ISO-THES e UNESKOS. Assim, esse segundo *corpus* de análise integrou 10 publicações, sendo que 5 delas já constavam no primeiro *corpus*. Contudo, ocorreu que uma referência, que não constou em nenhum dos resultados e recortes, foi bastante citada e destacada em parte das publicações analisadas e, outras duas referências foram identificadas como fundamentais para a discussão, sendo assim, optou-se pela seleção e leitura destas. Portanto, no total, foram analisadas 25 publicações, incluindo a norma ISO 25964 (partes 1 e 2) e recomendações SKOS.

A quarta etapa consistiu na leitura integral das publicações com a finalidade de identificar as similaridades entre as recomendações do modelo de dados SKOS e as recomendações da norma ISO 25964 para a representação de tesouros. A identificação e análise de tais similaridades permitiu ordenar as abordagens e aplicações, que envolvem a relação entre a norma ISO 25964 e o modelo de dados SKOS, em três grupos: a) Fundamentação histórico-conceitual; b) Técnicas de implementação, manutenção e gerenciamento de tesouros; e c) Recomendações, normalização, extensões e usos. Neste contexto, a expressão “abordagens” se refere aos aspectos de assunto aos quais o conteúdo da publicação mais se aproxima, enquanto a expressão “aplicações” remete às possibilidades de utilização prática do conteúdo da publicação.

A partir da composição desses três grupos, foi possível chegar à quinta etapa: concluir a análise de conteúdo categorial propriamente dita, de modo a identificar e discutir as seguintes categorias elementares de similaridades e equivalências entre o modelo SKOS e a norma ISO 25964 para construção de

tesauros: definição de tesauros com relação a outros SOC; conceitos e termos; domínios, microtesauro e grupos de conceitos; relações de equivalência; relações hierárquicas; principais termos e principais conceitos; relações associativas; notas; arranjos; equivalência composta; e extensões SKOS e outras propostas de aprimoramento do SKOS em atendimento à norma ISO 25964. Tais categorias foram delineadas conforme os pontos de discussão que apareciam com mais frequência nos textos das 25 publicações analisadas nos resultados, principalmente nos artigos científicos recuperados nas bases de dados nacionais e internacionais.

Os resultados da referida análise são apresentados e discutidos na seção 5, no alcance do objetivo de apresentar as similaridades entre o modelo SKOS e a norma ISO 25964 para a representação de tesauros, a partir da literatura científica.

A partir do exposto, torna-se válido ressaltar o panorama geral das 627 publicações inicialmente encontradas. Todas elas também se enquadraram em três grupos ou categorias de abordagens e aplicações da representação de diversos SOC em SKOS: a) Fundamentação histórico-conceitual; b) Técnicas de implementação, manutenção e gerenciamento de tesauros; e c) Recomendações, normalização, extensões e usos. E no âmbito das 627 publicações referentes a SKOS de um modo geral, os SOC do tipo tesouro foram os que mais se destacaram, integrando cerca de 200 publicações.

### **3 ENCADEAMENTOS ENTRE CIÊNCIA DA INFORMAÇÃO E TESAUROS**

Há uma história que precede e fundamenta os estudos e aplicações atuais que objetivam adaptar os tesauros para ambientes semânticos da Web. Portanto, nesta Seção, com o propósito de embasar a relevância de representar tesauros em SKOS, os principais conceitos e influências acerca da recuperação da informação são apresentados em consonância com a Ciência da Informação enquanto campo científico e profissional. Na subseção 3.1 estão identificadas as principais correntes epistemológicas da Ciência da Informação. Na subseção 3.2 são apresentadas as relações entre linguagem, informação, representação temática da informação e recuperação da informação. Por fim, na subseção 3.3, o tesauro recebe tripla caracterização, sendo conceituado como um instrumento terminológico, um sistema de organização do conhecimento e uma linguagem documentária.

#### **3.1 Um percurso epistemológico em Ciência da Informação**

A Ciência da Informação emerge como resultado dos desafios enfrentados por pesquisadores norte-americanos no campo da recuperação da informação. Esses pesquisadores se depararam com o aumento do volume de conhecimento científico publicado, o que os levou a desenvolver mecanismos e estratégias para garantir a organização e a recuperação eficiente dessas informações. Esses mecanismos e estratégias foram incorporados em sistemas e bases de dados automatizados, permitindo a gestão mais eficaz do conhecimento científico (Saracevic, 1999; Le Coadic, 2004).

Como campo de estudo, a Ciência da Informação surgiu após a Segunda Guerra Mundial, juntamente com outros campos como a Ciência da Computação, num cenário de intensas mudanças econômicas e industriais que propiciaram o surgimento de um fenômeno denominado “explosão da informação”, que apresentou um grande crescimento de publicações técnicas e científicas, agregando ainda mais valor competitivo à informação (Saracevic, 1999).

A Ciência da Informação direciona-se para a cientificidade, estabelecendo práticas para o gerenciamento eficiente da informação e do conhecimento em diversos setores da sociedade, abrangendo várias instituições e organizações. Para

isso, fundamenta-se em diferentes paradigmas, com o objetivo de facilitar o uso da informação por parte dos seres humanos. Nesse sentido, as tecnologias da informação e comunicação, juntamente com o usuário da informação, desempenham um papel dominante nos paradigmas da área (Le Coadic, 2004).

A Ciência da Informação se constitui em um campo científico, pois, na perspectiva moderna, a atuação de seus pesquisadores é concretizada a partir de estudos assentados “em duas ideias nucleares: a distinção entre o sujeito e o objeto e a produção de conhecimento disciplinada pelo método [...] como protagonista da neutralização da complexidade” (Tálamo; Smit, 2007, p. 32, tradução livre). No entanto, numa perspectiva pós-moderna, a Ciência da Informação propõe “a elaboração de conhecimento [...] determinado por temáticas” (Tálamo; Smit, 2007, p. 34, tradução livre). As perspectivas moderna e pós-moderna de Ciência da Informação influenciam na contemporaneidade do campo.

Para Saracevic (1999), a Ciência da Informação, como ciência e como campo profissional, é definida pelos problemas que abordam e os métodos que são utilizados para a resolução desses problemas. O autor ressalta três características evolutivas e existenciais da Ciência da Informação. A primeira é que a Ciência da Informação é interdisciplinar e as relações com outras disciplinas (como a Biblioteconomia, Ciência da Computação, Linguística e Educação) estão em constante mudança e longe de terminar. A segunda é que a Ciência da Informação é conectada com a tecnologia e que diante disto há um imperativo tecnológico que está restringindo a evolução da Ciência da Informação, assim como a evolução de outros campos e da sociedade da informação de um modo geral. E a terceira característica é que a Ciência da Informação está inserida na evolução da sociedade da informação e contém uma forte dimensão social e humana que vai além da tecnologia. Através destas três características, as concepções de recuperação da informação (prover e processar a informação com base na lógica), relevância (associar os processos às necessidades de informação das pessoas) e interação (permitir a troca e *feedbacks* entre sistemas e usuários que se utilizam de processos de recuperação da informação) integram as bases para compreensão da Ciência da Informação (Saracevic, 1999).

Assim, o termo Ciência da Informação aparece a partir de uma crise da visão racionalista do mundo e para que essa aparição seja melhor entendida, é preciso retomar alguns pontos das propostas teórico pragmáticas de quatro importantes teóricos da área: Gabriel Naudé, Melvil Dewey, Paul Otlet e De Solla Price. Para Naudé (1600-1653), a liberdade humana é possível a partir do acesso informacional irrestrito, comparação criteriosa e livre de preconceitos entre diferentes fontes de informações e opiniões (Tálamo; Smit, 2007). Melvil Dewey (1851-1931) direcionou seus trabalhos para concretizar a Biblioteconomia enquanto profissão “obtida por duas vias: ensino e associações profissionais” (Tálamo; Smit, 2007, p. 43). Com sua bagagem em organização e processos de referência, Dewey contribuiu para que a origem da Ciência da Informação também se fundamentasse no conceito de Biblioteconomia Moderna “associada aos processos de criação das coleções e aos modos de transformá-las em serviços” (Tálamo; Smit, 2007, p. 43). Derek de Solla Price (1922-1983), conhecido como o “pai da cienciometria”, afirma que a Ciência da Informação segue os princípios da racionalidade moderna. Paul Otlet (1968-1944) “previu as redes de informação e imaginou um sistema ágil e dinâmico que lhe permitisse interconectar as informações de acordo com a necessidade” (Tálamo; Smit, 2007, p. 45), conforme apresentado a seguir.

O conceito de Ciência da Informação também tem origens na Documentação e começou a ser utilizado por volta do ano de 1955 no âmbito dos trabalhos encabeçados pelos advogados belgas Paul Otlet e Henri La Fontaine no Instituto Internacional de Bibliografia (IIB), posteriormente denominado Federação Internacional de Documentação (FID) (Capurro, 2003; Rayward, 1997). O IIB teve propósitos ajustados a partir das motivações de Otlet em estabelecer formas de extrair sistematicamente o conhecimento, classificar e relacionar os assuntos (Rayward, 1997). Para tanto, o apoio político de La Fontaine foi fundamental para a criação da Classificação Decimal Universal (CDU) como uma das principais contribuições da parceria entre os dois advogados. O IIB detinha *databases* e coleções, e os bancos de dados catalográficos analógicos continham dados bibliográficos internacionais padronizados em fichas catalográficas. As coleções eram mapeadas, interrelacionadas e sistematizadas em benefício da recuperação de documentos. Além dessa estrutura e organização, o IIB contava com serviços de busca e disseminação da informação que permitiam o uso de expressões analógicas

de busca e disponibilização de materiais bibliográficos; tudo feito de forma manual (Rayward, 1997).

Paul Otlet, considerado o pai da Documentação, propagou ideias muito importantes para a fundamentação da Ciência da Informação até os dias atuais (Smit, 2008). O princípio monográfico é uma dessas ideias, definido como síntese que reúne informações sobre o mesmo tema, “tema único”, gerando, assim, um novo documento sintetizador (Smit, 2008), configurando-se como uso de tecnologia de cartão ou ficha que continha gravações de elementos principais de informações mais complexas (Rayward, 1997). A concepção de consulta como uma interrogação intencional de fontes de informação em busca de “bits” necessários demonstra que Otlet detinha uma visão diferenciada dos padrões de serviço de referência da época, pois a enfocava no desenvolvimento de mecanismos para o acesso efetivo à informação e não apenas no armazenamento (Rayward, 1997). Otlet também defendia a ideia de que o trabalho final da documentação seria uma codificação enciclopédica que ligaria informações contidas em materiais e elementos dispersos em publicações relevantes (Rayward, 1997; Smit, 2008). A ideia de uma rede universal para documentação se aproximou muito da concepção de internet, porém sua concretização dependia de novas tecnologias para a organização sistemática do trabalho documental (Otlet, 1937; Rayward, 1997). O nível mais básico para existência dessa rede seria a integração de trabalho cooperativo e padronização entre indivíduos e instituições, tais como: escritórios de documentação, bibliotecas, arquivos e museus. O trabalho e as visões de Paul Otlet se posicionaram a favor da garantia de acesso universal ao conhecimento como forma de garantir a paz mundial, de modo a promover união de todos os povos para chegar a uma pacificação por meio da difusão do conhecimento (Otlet, 1937). Para ele, o livro não era o único recurso informacional a ser considerado como documento, ponderando, inclusive, sobre a possibilidade de junção entre palavra falada e documentos a partir da utilização de microfilme (Rayward, 1997). Mesmo ainda não existindo a gama atual de TIC, a ideia de realidade virtual difundida por Otlet consistia em pensar que em algum lugar do mundo seria erguido um edifício contendo todos os livros e outras categorias de recursos informacionais e, nas mesas de trabalho não haveria livros ou outros documentos, mas apenas uma tela e um telefone para acessar as informações deste edifício (Rayward, 1997). Por considerar que a mente humana é

falha, Paul Otlet também propagou a ideia de um “livro universal” derivado de todos os livros existentes, que funcionaria como um tipo de anexo ao cérebro, um substrato da memória atrelado à existência de “uma máquina para explorar o tempo e o espaço” (Otlet, 1903, p. 86). Estas ideias de Paul Otlet precedem e combinam com a idealização do Memex, máquina imaginada por Vannevar Bush que possibilitaria o acesso informacional e navegação por entre vários textos por meio de conexões via “*hiperlinks*” (Bush, 1945), o que remete a ideia de computador e internet.

Na visão de Otlet, a Documentação, em relação à Bibliografia, carrega uma abordagem mais ampla para organização do conhecimento, porque considera fontes diversas (como imagens, textos, microfilmes e mapas), não servindo apenas para organizar documentos, como previsto por algumas bibliotecas na época, mas também para elaborar formas de extrair (como mapas mentais e conceituais) e organizar informações contidas nos diferentes suportes informacionais em prol da recuperação e consulta à informação (Rayward, 1997). Embora as propostas e trabalhos de Otlet não se refiram apenas a tarefas, pois contém teorias que alicerçaram suas propostas, suas publicações não se remontam às estruturas de evidências científicas (Rayward, 1997). Ainda assim, a Documentação está englobada pela Ciência da Informação e seus preceitos são fundamentais para o entendimento desta (Rayward, 1997).

Nesse sentido, a reflexão epistemológica a partir da Ciência da Informação é fundamental para a composição dos polos epistemológico e teórico do modelo metodológico sob o qual percorre esta dissertação. A epistemologia, que pode ser entendida como “estudo crítico dos princípios, das hipóteses e dos resultados das diversas ciências” (Tesser, 1995, p. 92), para Capurro (2003), se relaciona com a Ciência da Informação ao longo de uma complexa história em que é necessário delimitar o uso e a definição do conceito de informação no âmbito desta ciência. Deste modo, neste trabalho, o conceito de informação é trabalhado mais adiante, após a conceituação da Ciência da Informação.

Capurro (2003) destaca as correntes epistemológicas que influenciaram e ainda influenciam a Ciência da Informação, a citar: Hermenêutica, Racionalismo Crítico, Semiótica, Linguística e Pragmática, Construtivismo, Teoria da Informação,

Cibernética de segunda ordem e Teoria de Sistemas. Além dessas correntes, Capurro (2003) também apresenta as raízes da Ciência da Informação e se aprofunda nos três paradigmas epistemológicos deste campo científico: físico, cognitivo e social. As referidas raízes são a Biblioteconomia Clássica e a Computação Digital. A primeira pode ser entendida como “o estudo dos problemas relacionados com a transmissão de mensagens” (Capurro, 2003, p. [1]), enquanto a segunda

[...] é de caráter tecnológico recente e se refere ao impacto da computação nos processos de produção, coleta, organização, interpretação, armazenagem, recuperação, disseminação, transformação e uso da informação, e em especial da informação científica registrada em documentos impressos (Capurro, 2003, p. [1]).

O impacto com relação aos documentos impressos remete ao primeiro paradigma da Ciência da Informação, predominante entre 1945 e 1960, denominado paradigma físico e predominantemente influenciado pelas correntes epistemológicas Teoria da Informação, de Shannon e Weaver, e Cibernética de segunda ordem (Almeida, 2007; Capurro, 2003). Esse paradigma enfatizava a importância da existência da mensagem enquanto objeto em detrimento do papel do sujeito cognoscente que se apropria do conteúdo semântico da mensagem.

Na década de 70 ocorre um deslocamento para o paradigma cognitivo da Ciência da Informação, que se embasa na corrente epistemológica denominada Racionalismo Crítico, proposta por Karl Popper (Almeida, 2007; Capurro, 2003). Neste paradigma, destaca-se o enfoque no sujeito cognoscente atrelado a uma perspectiva objetiva da informação obtida de modo individualista e isolacionista dos processos sociais informativos (Capurro, 2003). Capurro (2003) cita as críticas de Frohmann ao paradigma cognitivo fundamentadas na epistemologia Linguística e Pragmática de Wittgenstein e na Teoria do Discurso como manifestação de poder de Michel Foucault, visto que “Bernd Frohmann [...] considera o paradigma cognitivo não só como idealista mas também como associal” (Capurro, 2003, p. [1]).

Os precursores do paradigma social são Hjørland e Albrechtsen (1995) que defendem “integração da perspectiva individualista e isolacionista do paradigma cognitivo dentro de um contexto social no qual diferentes comunidades desenvolvem seus critérios de seleção e relevância” da informação (Capurro, 2003, p. [1]). Assim,

nesse paradigma, os processos cognitivos informacionais se relacionam diretamente às comunidades discursivas (Araújo, 2017). De acordo com esse paradigma, “o objeto da ciência da informação é o estudo das relações entre os discursos, áreas de conhecimento e documentos em relação às possíveis perspectivas ou pontos de acesso de distintas comunidades de usuários” (Hjørland, 2003 *apud* Capurro, 2003, p. [1]).

A corrente epistemológica hermenêutica influencia o paradigma social a medida em que

A hermenêutica como paradigma da ciência da informação postula justamente a diferença entre pré-compreensão, oferta de sentido e seleção, tomando como marco de referência, não a pré-compreensão de um sujeito ou usuário isolado, mas a de determinada comunidade assim como a de um campo específico de conhecimento e/ou de ação no qual o usuário está já implícita ou explicitamente inserido. Nesse sentido, o paradigma hermenêutico está próximo da semiótica, assim como do construtivismo e da cibernética de segunda ordem (Capurro, 2003, p. [1]).

Nesse sentido, os domínios ou áreas do conhecimento podem ser entendidos como resultados da ação de grupos ou indivíduos sob determinados temas que interessam aos fazeres demandados por estas comunidades. Assim, “os termos de um léxico não são algo definitivamente fixo, [pois o] trabalho informativo é um trabalho de contextualizar ou recontextualizar praticamente o conhecimento” (Capurro, 2003, p. [1]). “Informação é conhecimento em ação” (Kuhlen, 1996, p. 34 *apud* Capurro, 2003, p. [1]).

Capurro (2003) afirma que a existência desses três paradigmas não se enquadra numa sequência linear, pois em alguns momentos da história existiram posicionamentos intermediários de pensamento e prática por entre esses paradigmas da Ciência da Informação. Assim como os paradigmas físico e cognitivo, o paradigma social também recebe críticas a partir de suas limitações, pois no âmbito da Ciência da informação, a discussão sobre o conceito de informação, ora definida a partir de processos cognitivos (conhecimento tácito) e ora como conhecimento registrado, pressupõe limites a todo paradigma ou modelo (Capurro, 2003).

nesse caso do paradigma social, **no momento em que a relação entre informação e significado torna-se problemática quando se deseja transportá-la para sistemas não sociais.** É aqui que surge

o apelo por uma teoria unificada da informação (Hofkirchner, 1999). Essa teoria deveria entrecruzar ou, por assim dizer, enredar ou tramar diversos conceitos de informação **mostrando a tessitura complexa da linguagem comum** e da teorização científica em torno desse conceito e a sua relação com a realidade social e natural que o possibilita (Capurro; Hjørland 2003; Capurro 2001a, *apud* Capurro, 2003, p. [1], grifo nosso).

Os sistemas de recuperação da informação estão entre os sistemas não sociais para os quais intenciona-se transportar informação e seus significados. Por conseguinte, recortes acerca da teorização dos conceitos de recuperação da informação, informação e linguagem são apresentados.

### **3.2 Linguagem, informação, representação temática da informação e recuperação da informação**

A linguagem é mecanismo primordial para construção e comunicação da informação, manifestando-se à medida em que simultaneamente modifica e é modificada pelas diferentes necessidades de atribuição de sentidos a serem declarados nos mais diversos contextos de ação. Diante das ações que configuram os processos de organização do conhecimento, a linguagem é um instrumento para acessar e se apropriar da informação de modo a produzir novo conhecimento. Moreira (2012, p. 125) aponta que

Os problemas relacionados à obtenção de informação nos ‘novos’ sistemas de informação são problemas de linguagem [...fazendo com que] a questão terminológica se aproxime da documentação, cujos processos e produtos serão estudados da Linguística Documental.

No tangente à Linguística, a linguagem e a língua se relacionam a partir da necessidade de atribuição de significação. A linguagem é a capacidade que os seres humanos têm para produzir, desenvolver e compreender a língua e outras manifestações, como a pintura, a música e a dança (Petter, 2010). Por sua vez, a língua, conforme a Linguística Estrutural de Saussure, é estruturada numa rede de relações que forma um sistema de signos que condicionam a significação através da comunicação por meio de LN ou uma LD (Cintra *et. al*, 2002; Silva Filho, 2020). Nesse sentido, no âmbito da Semiótica cunhada por estudos de Peirce (1995, p. 45), o signo “ou *representamen*, é aquilo que, sob certo aspecto ou modo, representa algo para alguém”. Considerando a significação, tem-se que a semiose é a produção de significados através de um processo pelo qual qualquer coisa age como um signo

e, por agir assim, significa, ou seja, atribui semântica (Noth, 1995, 2005; Peirce, 1995).

Atentando-se para as influências que a Ciência da Informação recebe da epistemologia Linguística e Pragmática de Wittgenstein, do contexto de produção e uso da informação a partir do paradigma social, da concepção de informação enquanto conhecimento em ação e da constatação da volatilidade dos termos de um léxico, é plausível confluir que

A linguagem está onde a informação está e onde a sociedade está. A compreensão sobre a relação sujeito–mundo (mediados pela tecnologia ou não) é fenomenológica, cibernética, psicológica, ecológica e transformadora. E a Pragmática tem sido o movimento filosófico que coloca a linguagem em uso, pelos sujeitos em sociedade, como o elemento articulador de todos os demais processos que envolvem os fatos e os atos no mundo. **A própria ação é uma linguagem e a linguagem, uma ação.** (Gracioso, 2016, grifo nosso).

Esta compreensão de linguagem corrobora com estudos em Filosofia da Linguagem em Ciência da Informação, na qual Gracioso e Saldanha (2011, p. 65) apontam que Wittgenstein defende que “a linguagem não poderia ser entendida a partir da lógica – ela precisaria ser entendida a partir de seu uso”. Neste sentido, a linguagem existe a partir do contexto comunicacional onde os discursos são elaborados pela necessidade de produção de sentidos em benefício do registro de novos conhecimentos, do entretenimento, realização de tarefas e outros afazeres, dos mais simples aos mais complexos, interessantes para a manutenção da qualidade de vida do indivíduo e da vida em sociedade. Assim, considera-se que cada campo ou domínio do conhecimento produz informações formuladas de acordo com a linguagem peculiarmente empregada nas dinâmicas de comunicação dos atores do domínio. Esta consideração se fundamenta, também, no fato de que “Wittgenstein proporia que, para que se consiga compreender a linguagem, suas funções práticas precisariam ser entendidas, e não somente o significado isolado das expressões linguísticas” (Gracioso; Saldanha, 2011, p. 66).

Portanto, pensar a representação temática da informação em qualquer ambiente, em particular o digital, é conjecturar sobre as dinâmicas de emissão e apropriação da informação para propor sistemas de ORC suscetíveis de constante manutenção em prol de dinâmica recuperação da informação. Como formulou

Ranganathan (2009), a biblioteca é um organismo em crescimento, e assim é a Web, que não cresce como uma biblioteca, pois não o é, também não cresce apenas em quantidade de dados e informações, mas também se expande em novos conceitos (ideias) e significados emergidos das transformações ativas da linguagem.

Deste modo, o conhecimento registrado se constitui em informação que é definida a partir de contextos específicos, mas sempre como “um conhecimento inscrito (registrado) em forma escrita (impressa ou numérica [digital]), oral ou audiovisual, em um suporte, o qual comporta um elemento de sentido” (LE COADIC, 2004, p. 5). A informação, então, possui características intrínsecas à sua composição ontológica, uma vez que pode ser registrada (codificada) de diversas formas; duplicada e reproduzida *ad infinitum*; transmitida por diversos meios; conservada e armazenada em suportes diversos; medida e quantificada; adicionada a outras informações; organizada, processada e reorganizada segundo diversos critérios e recuperada quando necessário, segundo regras pré-estabelecidas (Robredo, 2003).

No cerne da comunicação, a informação é analisada levando em consideração os usuários como o foco principal, uma vez que o processo comunicativo só é completo quando a mensagem é processada pelo receptor. No processo de transmissão de mensagens (comunicação), são considerados os usuários envolvidos, o canal de comunicação formado por meio de artefatos tecnológicos, bem como a informação em si, contida no conteúdo das mensagens (Le Coadic, 2004).

Ao caracterizar o que é informação na Ciência da Informação, Saracevic (1999, p. 1054, tradução livre) afirma que “a informação é um fenômeno básico” e que, como todos os fenômenos básicos, não é algo tão simples de definir. Em sentido específico, “a informação é considerada em termos de **sinais ou mensagens** para decisões envolvendo pouco ou nenhum processamento cognitivo, ou tal processamento que pode ser expresso em algoritmos e probabilidades” (Saracevic, 1999, p. 1054, tradução livre, grifo nosso). Em sentido amplo, a informação se configura ao envolver “diretamente o **processamento cognitivo** e a compreensão. Resulta da interação de duas estruturas cognitivas, uma "mente" e (amplamente) um "texto". Informação é o que afeta ou muda o estado de uma

mente” (Saracevic, 1999, p. 1054, tradução livre, grifo nosso). Porém, em sentido mais amplo,

**A informação é tratada em um contexto.** Ou seja, a informação envolve não apenas mensagens (primeiro sentido) que são processadas cognitivamente (segundo sentido), mas também um contexto - situação, tarefa, problema em questão e assim por diante. Usar informações que foram processadas cognitivamente para uma determinada tarefa é um exemplo. Além de outros sentidos, **envolve motivação ou intencionalidade e, portanto, está ligada ao contexto ou horizonte social expansivo, como cultura, trabalho ou mão-problema** (Saracevic, 1999, p. 1054, tradução livre, grifo nosso).

Diante desses três sentidos que colaboram para o entendimento sobre o que seja informação, percebe-se que o *sentido específico*, o *sentido amplo* e *sentido mais amplo* se relacionam respectivamente com os três paradigmas da Ciência da Informação defendidos por Capurro (2003): o paradigma físico, o paradigma cognitivo e o paradigma social. Deste modo, para Saracevic (1999, p. 1054, tradução livre), “na ciência da informação, devemos considerar o terceiro e mais amplo sentido de informação, porque a informação é usada em um contexto e em relação a alguns motivos”, em que a concepção de informação no sentido mais amplo foi implicitamente compreendida desde o início da Ciência da Informação, principalmente perante as práticas de recuperação da informação (Saracevic, 1999).

No âmbito da Ciência da Informação, um elo teórico e prático entre linguagem e informação é a recuperação da informação. É através do uso da linguagem que se pode exercer a atividade de busca para recuperar informações que satisfaçam determinada necessidade de informação (como se o sujeito cognoscente se comunicasse com o sistema a partir da linguagem para poder declarar as informações das quais precisa). Para tanto, a linguagem em sua forma natural e a linguagem em sua forma artificial são fundamentais para a formulação de estratégias de busca de informação. Na primeira forma, tem-se que a linguagem natural é ambígua, sinonímica, de alta revocação e baixa precisão; enquanto que na segunda forma, cita-se a linguagem documentária, que por sua vez está submetida a padrões, normas, precisão terminológica, pertinência e serve como conversora e mediadora sistêmica entre a linguagem natural do usuário informacional e a linguagem contida nos assuntos abordados por um conjunto de documentos indexados num SRI (Campos, 2001; Cintra *et al.*, 2002; Toutain, 2007). Assim sendo,

um dos desafios detectados no âmbito da busca e recuperação da informação é o estabelecimento de interoperabilidade semântica entre LDs, humanos e máquinas (Barbosa; Viera; 2022; Pastor-Sánchez; Martínez-Méndez; Rodríguez-Muñoz, 2012).

Diante de tal desafio, a Ciência da Informação abarca recursos teóricos e metodológicos de disciplinas como a Linguística, Semiótica, Terminologia e Organização do Conhecimento (*Knowledge Organization* - KO) em prol da proposição de linguagens documentárias enquanto Sistemas de Organização do Conhecimento (*Knowledge Organization System* - KOS) que propiciem processos de organização do conhecimento (*Knowledge Organization Process* - KOP) adequados às demandas de RI (Hjørland, 2016). Nesta ótica,

A recuperação de informações é [...] o processo de localização ou descoberta em relação às informações armazenadas [...] útil para [um usuário]. A recuperação da informação abrange os aspectos intelectuais da descrição da informação e sua especificação para pesquisa, e também quaisquer sistemas, técnicas ou máquinas que são empregados para realizar a operação (Mooers, 1951 *apud* Saracevic, 2009, p. 5, tradução livre).

Com efeito, “um sistema de recuperação da informação é um ambiente linguístico cuja eficiência depende de um controle adequado da representação dos itens de informação e das requisições de seus usuários” (Ferneda; Dias, 2015, p. 1). Assim, a mediação da comunicação entre sujeitos cognoscentes e informações requisitadas ocorre por meio de infraestrutura tecnológica adequada que inclui representações temáticas (representar tematicamente o documento) e descritivas (descrever fisicamente o documento) de informações legíveis pelos mecanismos de busca (como algoritmos) capazes de revocar itens informacionais relevantes (úteis) aos interesses desses sujeitos (Ferneda; Dias, 2015; Saracevic, 2009).

A relevância das informações recuperadas num SRI, entre outros fatores concernentes à infraestrutura tecnológica e familiaridade de usos e usuários, atrela-se à qualidade da abordagem temática das informações representadas, além da representação descritiva. No contexto específico do tratamento temático da informação, são analisadas características intelectuais nos objetos de informação, permitindo, assim, a criação de registros representativos do assunto. Esses registros são manipulados durante a busca e recuperação como pontos de acesso informacional, sendo essenciais para a seleção e o acesso às informações (Foskett,

1973). Dessa forma, a informação contida num documento se torna acessível ao usuário que precisa dela, atendendo a um dos principais interesses da Ciência da Informação enquanto campo científico e profissional que, atrelada à Biblioteconomia, vem examinando o conceito de assunto há mais de um século (Hjørland, 2017). Porém, o tratamento temático enquanto contexto de dedicação processual ao assunto, carece de avanços metodológicos e instrumentais (Pando, 2018).

O tratamento temático da informação, também denominado de representação temática, análise documentária, análise temática, análise de assunto ou até mesmo catalogação de assunto, refere-se ao procedimento de síntese sobre o conteúdo de um documento, com o propósito de expor os atributos temáticos presentes neste documento (Foskett, 1973). O Tratamento Temático da Informação (TTI) é uma subárea de estudo da ORC.

Antony Charles Foskett impulsionou o termo Tratamento Temático da Informação a partir da década de 70, com a publicação da obra *The subject approach to information*, cuja tradução para o Brasil, feita por Antonio Agenor Briquet de Lemos, foi publicada em 1973. A referida obra aborda elementos ao entorno de: características de um sistema de recuperação de informações, o documento como evidência, análise de assuntos, cabeçalhos alfabéticos de assuntos, arranjo sistemático, esquemas gerais de classificação, notação, índice alfabético, tipos de índices pré-coordenados, linguagens de indexação pré-coordenada, sistemas pós-coordenados, linguagens de indexação pós-coordenada e perspectivas futuras relacionadas à pesquisa em recuperação da informação e ao computador. No que concerne a este último elemento, “é preciso que tenhamos a certeza de que os alimentamos [os computadores] com o conjunto correto de sinais aos quais devem responder, isto é, devemos alimentá-los com os dados corretos” (Foskett, 1973, p. 398). O autor ainda afirma que “a programação é uma técnica que normalmente não faz parte da formação do bibliotecário” (Foskett, 1973, p. 398), porém é necessário que saibamos analisar as necessidades computacionais da biblioteca “para que o programador possa utilizar a análise feita por nós na compilação de um programa [software] que alcance o objetivo colimado” (Foskett, 1973, p. 398), principalmente no que tange à recuperação temática da informação.

Os avanços na área de desenvolvimento de sistemas de recuperação da informação são progressivos desde o início da década de 50 do século XX, período de origem da Ciência da Informação. “No início, a RI era estática. Agora é altamente interativa” (Saracevic, 2009, p. 5, tradução livre). Além de representações em índices e resumos, os SRI passaram a lidar com textos completos e diversas mídias que não se restringem apenas à impressão e aos textos, chegando a cobrir cenários informacionais da Web (Saracevic, 2009).

A avaliação dos SRI também é fundamental e ocorre a partir de testes que analisam sua eficiência e eficácia mediante índices de precisão e revocação (Saracevic, 2009). Neste sentido, os níveis de especificidade e exaustividade da indexação também influenciam nesses índices (Lancaster, 2004), haja vista que

A precisão mede quantos itens recuperados [...] eram relevantes ou, inversamente, quantos eram ruídos. A revocação mede quantos dos itens potencialmente relevantes em um determinado arquivo ou sistema foram realmente recuperados ou, inversamente, quantos não foram recuperados, embora fossem relevantes (Saracevic, 2009, p. 7).

Todavia, além da importância dos modelos e sistemas de organização do conhecimento discutidos a partir da Biblioteconomia e Ciência da Informação, toda a infraestrutura que envolve os processos de RI em SRI compreende tecnologias e modelos advindos de campos como a Estatística, Matemática e Computação (Baeza-Yates; Ribeiro Neto, 2013; Coneglian, 2020). As próprias cinco leis da Biblioteconomia foram formuladas por um bibliotecário e matemático (Ranganathan, 2009). Os desafios ao entorno da evolução e demandas por melhorias dos SRI são diversos e interdisciplinares e podem estar centrados em dados e informações ou na interação com os usuários (Baeza-Yates; Ribeiro Neto, 2013). Com relação à Web, esses desafios ficam ainda mais complexos, pois envolvem uma diversidade de padrões, origens documentais e crescente quantidade de dados que dificultam a interoperabilidade e a encontrabilidade da informação<sup>5</sup>.

Desta perspectiva, é fundamental que a estrutura dos SOC, em especial os do tipo tesouro, estejam construídas para funcionar enquanto instrumentos

---

<sup>5</sup> Para saber mais sobre o conceito de encontrabilidade da informação, leia: VECHIATO, F. L. Encontrabilidade da informação: contributo para uma conceituação no campo da ciência da informação. 2013. 206 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013.

terminológicos que se relacionem operacionalmente com as demais ferramentas elaboradas a partir de outras áreas do conhecimento, mas que também são de suma importância para a composição de SRI.

### **3.3 Tesouro: instrumento terminológico, SOC e linguagem documentária**

A terminologia é um conjunto de termos de um domínio do conhecimento; envolve conceitos, sistema de conceitos e termos, tendo como fontes para construção de SOC os discursos da área de especialidade do domínio representado (Jesus; Moreira, 2018; Tálamo; Lara; Kobashi, 1992). Os discursos registrados de uma determinada ciência ou área de especialidade servem de base para identificar a terminologia do domínio do conhecimento que se pretende representar. Já a Terminologia, enquanto disciplina, abrange diferentes teorias terminológicas, tais como: Teoria Geral da Terminologia cunhada por Wüster, na Alemanha (Barros, 2004), em que o conceito independe da língua e o signo linguístico é instrumental, pois apenas denomina e ocorre a prescrição e padronização terminológica; Teoria Comunicativa da Terminologia difundida por Cabré (1995), na Espanha, para a qual o fenômeno terminológico ocorre do interior da linguagem, tendo a terminologia como descritiva; e a Socioterminologia, conceituada por Gaudin, na França (Faulstich, 2006), que considera o uso social dos termos e também julga a terminologia como descritiva.

SOC são instrumentos terminológicos que podem ser exemplificados em sistemas de classificação, listas de cabeçalho de assunto, glossários, vocabulários controlados, tesouros, ontologias, mapas conceituais e taxonomias. Estes instrumentos se configuram como operadores de sentido (semântica) ao passo que estruturam conceitos conforme domínios específicos do conhecimento e, por esta configuração, são compreendidos como instrumentos de representação temática da informação (Barbosa, 1969; Barité, 2011; Barros, 2004; Currás, 2005; Hjørland, 2016; Tálamo; Lara; Kobashi, 1992).

Barité (2011) aponta que do início da última década do século XX e durante a primeira década do século XXI, foram identificados dois processos concomitantes nos quais os instrumentos utilizados para representação e recuperação temática de documentos e recursos de informação têm evoluído e diversificado. Além do 1)

progresso no desenvolvimento de sistemas de classificação, tesauros e listas de cabeçalhos, 2) novas estruturas conceituais foram introduzidas, que estão relacionadas a aplicações específicas de *software*, como ontologias, ou a propósitos específicos dentro do fluxo de informações, como os anéis de sinônimos. Barité (2011, p. 126, tradução livre) também sugere a denominação Sistema de Organização do Conhecimento (SOC) “como um termo genérico abrangente para se referir ao conjunto de ferramentas destinadas à classificação e indexação”. O termo foi cunhado em espanhol pela professora espanhola San Segundo, na década de 90, antes mesmo da tradução para o inglês, *Knowledge Organization System* (KOS), início dos anos 2000. É neste sentido que a presente pesquisa se utiliza do termo Sistemas de Organização do Conhecimento (SOC), para se referir tanto às linguagens tradicionais criadas para indexação e classificação como para estruturas conceituais pensadas para outras finalidades, “que foram desenvolvidas como referência conceitual, enquanto estruturas semânticas, em qualquer área do conhecimento e, portanto, em sua documentação [, a saber:] mapas de tópicos, mapas conceituais e, de forma mais polêmica, até mesmo dicionários e glossários” (Barité, 2011, p. 126, tradução livre).

Os SOC são estudados por diferentes disciplinas, mas em especial pela KO. Para compreender a KO enquanto disciplina é preciso esclarecer que o conhecimento registrado, além de acessado por meio da linguagem, também se constitui a partir desta, haja vista que é representado por uma estrutura de conceitos conectados via relações semânticas identificadas a partir da comunicação do conhecimento (HJØRLAND, 2016). Estas relações precisam ser acionadas quando um indivíduo ou grupo se depara com a necessidade de aprender algo e demonstra uma defasagem de conhecimento a ser suplementada a partir de formulação de estratégia de busca para a recuperação da informação, em favor da aquisição e produção de novo conhecimento. “Aqui a informação é qualificada como instrumento modificador da consciência do homem. Quando adequadamente apropriada, produz conhecimento e modifica o estoque mental de saber do indivíduo” (Barreto, 2002, p. 70).

Para que a informação seja adequadamente apropriada, o conhecimento precisa estar organizado e sistematicamente representado a partir da gama de conceitos que o integram. O conceito, então, enquanto unidade do conhecimento

representado numa forma verbal (Dahlberg, 1978; Maculan; Lima, 2017) é elemento chave para a construção de LD enquanto um dos tipos de SOC. Neste sentido, “a organização do conhecimento em sistemas de classificação e sistemas de conceito são assuntos centrais em KO, afiliada principalmente à Biblioteconomia e Ciência da Informação e se constitui num campo de pesquisa, ensino e prática” (Hjørland, 2016, p. [1], tradução livre).

Da relação entre a KO, Linguística, Terminologia e a Documentação de origem francesa, pode-se citar a existência da Linguística Documentária difundida por García Gutiérrez, “que constitui um subcampo da Ciência da Informação que tem como objetivo estudar os problemas que caracterizam a Linguagem Documentária” (Lara, 2011, p. 114). A LD é uma linguagem artificial, sendo construída para intermediar a linguagem natural do usuário informacional e a linguagem dos documentos que potencialmente detêm a informação necessitada por ele; possui uma estrutura em que a significação ocorre por meio dos relacionamentos entre seus termos ou descritores contextualizados numa hipótese de organização que simula as conexões terminológicas usuais de determinado domínio do conhecimento (Campos, 2001; Cintra *et al.*, 2002; Currás, 2005; Lara, 2011; Smit, 2015). O objetivo da LD é a organização, representação e recuperação da informação. Na década de 50 e 60 surge a preocupação em torno da necessidade de desenvolver uma teorização das linguagens documentárias, ou seja, das metalinguagens específicas utilizadas nos ambientes de tratamento da informação. O que anteriormente era feito com base no senso comum advindo das práticas de representação temática passou a ser fundamentado por uma teoria sólida, fortemente embasada na linguística e, em certos contextos de estudo, com ênfase na terminologia, “visando questões relacionadas ao controle de vocabulário e ao poder comunicacional dos produtos da análise documentária” (Smit, 2015, p. 728).

A estrutura e composição de uma LD, segundo Gardin (1966, 1970, 1973), compreende um léxico, relações paradigmáticas e relações sintagmáticas. O léxico é a composição de “uma lista de elementos descritores e não descritores” (Berti Junior *et al.*, 2017, p. 383). As relações paradigmáticas são “definidas *a priori*, para indicar relações essenciais e, geralmente, estáveis entre os descritores” (Berti Junior *et al.*, 2017, p. 383), enquanto as relações sintagmáticas são definidas *a posteriori*, cunhadas para “determinar as relações contingentes entre descritores, válidas

apenas no contexto particular de uso” (Berti Junior *et al.*, 2017, p. 383). O termo LD é proposto por Gardin na década de 60, identificado como uma linguagem de indexação formada por “um conjunto de termos, providos ou não de regras sintáticas, utilizadas para representar conteúdos de documentos técnico-científicos com fins de classificação ou busca retrospectiva de informações” (Gardin *et al.*, 1968 *apud* Cintra *et al.*, 2002, p. 25). Assim, as LD são um tipo de SOC suscetíveis de serem utilizadas em SRI de diferentes unidades de informação e instituições de primeiro, segundo e terceiro setor.

Além de conhecimentos relacionados à Linguística, a construção de um SOC exige conhecimentos específicos sobre padrões, normas, formatos, contribuições da Computação e da Lógica (Campos, 2001). As exigências atreladas ao campo da Computação e Lógica ficam ainda mais evidentes diante da necessidade de pesquisa, construção, elaboração e uso de SOC em ambientes semânticos da Web e LOD.

As LDs se constituem em instrumentos imprescindíveis para a organização e representação temática de documentos de um acervo, visando a aproximação e recuperação por assunto (Barbosa, 1969; Currás, 1995). No tangível a exemplos de LD destaca-se que os tesouros e os sistemas de classificação documentária são as LD mais conhecidas e empregadas. A principal diferença entre esses dois tipos de LD está no grau de reprodução das relações presentes na LN e no escopo do conhecimento que pretendem abranger. Os primeiros sistemas de classificação bibliográfica conhecidos – *Dewey Decimal Classification* (CDD), *Classificação Decimal Universal* (CDU) e *Library of Congress* (LC) –, são de natureza enciclopédica e têm o objetivo de abranger todo o espectro do conhecimento. Por outro lado, os sistemas de classificação facetada, desenvolvidos posteriormente pelo *Classification Research Group* (CRG) com base na *Colon Classification* de Ranganathan, são voltados para domínios específicos. Os tesouros, por sua vez, surgiram a partir dessas classificações facetadas com uma preocupação adicional, que é o controle do vocabulário (Cintra, 2002).

O tesouro é identificado como uma dentre outros tipos de LD enquanto linguagens artificiais que são convencionalmente reconhecidas como “instrumentos de controle terminológico que atuam em dois níveis: a) na representação da

informação obtida pela análise e síntese de textos; b) na formulação de equações de busca da informação” (Tálamo *et al.*, 1992, p. 197). Para esta pesquisa, a LD considerada foi o tesouro pelo fato deste compreender um conjunto controlado de termos que representam conceitos significativamente relacionados de uma determinada terminologia, proporcionando a tradução entre LD e LN no âmbito de um sistema de recuperação da informação (Barros, 2004; Cabré, 1993, 1999; Currás, 2005; Tálamo *et al.*, 1992).

Os tesouros são concebidos como LD modernas, que contemplam basicamente três tipos de relações entre os termos (Cintra, 2002, p. 32): relação hierárquica, relação de equivalência e relação associativa, em que “organização hierárquica fundamental do tesouro funciona de forma dicionarial. Ela pressupõe uma definição do tipo ‘se > então” (Tálamo *et al.*, 1992, p. 198), baseada em um conjunto de regras e normas institucionais, sua extensão é determinada pelos acordos sociais, pressupondo modelos socioculturais e ideológicos específicos (Tálamo *et al.*, 1992). “O tesouro promove o controle de sinônimos e estruturas sintáticas simplificadas” (Dodebei, 2002, p. 66), de modo que direciona o processo de indexação e auxilia no processo de busca e ampliação de consultas. Atua, portanto, como ferramenta terminológica elaborada para a recuperação da informação (Barbosa; Viera, 2022).

Foskett (1973) apresenta o tesouro com linguagem de indexação pós-coordenada, apontando como um dos exemplos o “Thesaurus da Nasa” (Foskett, 1973, p. 350), onde os termos de busca são utilizados para formular estratégias de busca após a indexação dos documentos pelos mesmos termos. O autor aponta que “das linguagens de indexação preparadas especificamente para indexação pós-coordenada, as da ciência e tecnologia foram as primeiras” (Foskett, 1973, p. 337) sendo, na época, preponderantes, tanto em número quanto em áreas abrangidas. O autor elucida sobre a questão do mau uso do termo *thesauri* e *thesaurus* para designar listas de cabeçalho de assuntos e menciona o *Thesaurus of Engineering and Scientific Terms* (TEST), publicado em 1967, reconhecido como primeiro tesouro a influenciar a padronização e adoção de notações como BT (*Broader Term*), NT (*Narrower Term*) e RT (*Related Term*) para termo genérico, termo específico e termo relacionado respectivamente (Bandyopadhyay; Mukhopadhyay, 2015; Clarke; Zeng, 2012).

Barité (2011) apresenta os seguintes SOC: sistemas de classificação, códigos de classificação, listas de cabeçalhos de assuntos ou de epígrafes, tesouros, lista de descritores, lista de autoridades, anéis de sinônimo, taxonomia, ontologias, folksonomias, mapas conceituais, mapas de tópicos ou 'topic maps' e diretórios de buscadores. Dentre eles, “os tesouros constituem a ferramenta mais refinada de representação do conhecimento que foi criada até o presente momento” (Barité, 2011, p. 130), que mesmo sendo compreendido como a linguagem documentária mais bem elaborada e possuir uma estrutura que tem se mantido relativamente estável ao longo das décadas, novas ideias e perspectivas surgem continuamente, o que leva ao estabelecimento de regras específicas para facilitar a criação e a gestão de tipos particulares em situações específicas de uso, “como tesouros de usuário, tesouros de software, tesouros corporativos e outros” (Barité, 2011, p. 123).

Como exemplo da organização de um tesouro clássico, cita-se o Tesouro de Folclore e Cultura Popular Brasileira<sup>6</sup>, com parte representada na Figura 2:

**Figura 2** - Interface online do Tesouro de Folclore e Cultura Popular Brasileira.



Fonte: Tesouro de Folclore e Cultura Popular Brasileira (s. d.).

Analisando a Figura 2 é possível detectar as relações hierárquicas e associativas entre os termos do tesouro. As relações hierárquicas são divididas em relações genéricas (gênero-espécie) e relações partitivas (todo-parte) (Currás,

<sup>6</sup> Disponível em: <http://www.cnfcp.gov.br/tesouro/>

2005). Por exemplo, pode-se detectar a relação genérica entre o termo Dança e os Termos Específicos (TE) Anu, Arara (dança), Arurana, Baio, Baião e os demais tipos de dança representados terminologicamente no referido Tesouro. Um exemplo de relação partitiva, nesse contexto, é representado por Bananeira como Termo Genérico Partitivo (TGP) e caule, folhas e banana (fruto), como Termos Específicos Partitivos (TEPs) (Currás, 2005). As relações associativas, no caso, são estabelecidas entre Termo Genérico (TG) Ciranda (dança) e os Termos Associados (TA) – estes também poderiam denominar Termos Relacionados (TRs) – Ciranda (baile) e Dança de roda.

No contexto dos tesouros, os termos são representações de conceitos, estes que por sua vez são concebidos como ideias advindas das relações humanas com o mundo (Sowa, 1984). As siglas TGP, TEP, TG, TA e TR são exemplos de notações que um tesouro deve possuir para designar as relações semânticas entre os termos (Currás, 2005; Rautenberg; Souza; Kelnar, 2020). Seguem outras notações padrão para termos de um tesouro, conforme descritas no Quadro 3:

**Quadro 3** - Função de notações terminológicas de tesouros clássicos.

Notação terminológica	Significado	Tipo de relação indicada	Função
<b>TG</b>	Termo Genérico	Relação hierárquica	de sentido inverso para “Termo Específico”. Entre dois conceitos, indica aquele com conotação mais ampla. Por exemplo: o conceito “tesouro” tem o TG “Linguagens Documentárias”
<b>TE</b>	Termo Específico	Relação hierárquica	de sentido inverso para “Termo Genérico”. Entre dois conceitos, indica aquele com conotação mais restritiva. Por exemplo: o conceito “tesouro” tem com um TE “microtesouro”
<b>TR</b>	Termo Relacionado	Relação associativa	bidirecional entre dois conceitos. Por exemplo: “tesouro” é um TR a “sistemas de classificação”
<b>UP</b>	Usado Para	Relação de equivalência	de sentido inverso para “USE”. Indica a existência de outro termo sinônimo, o qual não é o

			preferido/adequado para rotular o conceito em análise. Por exemplo: o termo UP de “tecnologias da informação” é “tecnologia da informação e comunicação”
<b>USE</b>	Use	Relação de equivalência	de sentido inverso para “Usado Para”. Indica a preferência de uso para rotular o conceito. Por exemplo: “tecnologias da informação e comunicação” é USE para “tecnologia da informação”

Fonte: Elaborado a partir de Pinheiro e Ferrez (2014) e Rautenberg, Souza e Kelniar (2020).

Contudo, o Tesouro de Folclore e Cultura Popular Brasileira não está representado de forma interoperável semanticamente de acordo com recomendações internacionais para composição da Web Semântica. Haja vista que, “a interoperabilidade está associada à capacidade de mapear conceitos correspondentes em diferentes tesouros e demais vocabulários em dois níveis: lexical e semânticos” (Barbosa; Viera, 2022, p. 238). De um modo geral, num ambiente de conexão de dados, a interoperabilidade é a capacidade de troca e compartilhamento de informações a partir de características sintáticas e estruturais, mas quando se trata de relacionamentos semânticos, isto é, aqueles estabelecidos entre conceitos representados por termos de um tesouro, a interoperabilidade é semântica.

Quando se trata de interoperabilidade semântica, a atenção se volta não apenas para o intercâmbio, mas, principalmente, para o significado desses dados em diferentes sistemas de informação. No âmbito dos tesouros, a interoperabilidade é viabilizada quando os vínculos são estabelecidos entre conceitos de diferentes vocabulários. No contexto dos dados conectados, estabelecer esses relacionamentos com recursos externos, além de facilitar a identificação de correspondentes semânticos em conjuntos de dados heterogêneos, tem o potencial de estabelecer uma base lógica e conceitual consistente que ajuda os usuários a explorar e recuperar dados contextuais diretamente relacionados aos domínios representados no tesouro (Barbosa; Viera, 2022, p. 239).

Para estabelecer tal interoperabilidade, a definição de padrões, normas e modelos é crucial. Por isso, modelos de dados como o SKOS e a norma ISO

25964-1:2011<sup>7</sup> e ISO 25964-2:2013<sup>8</sup> devem ser considerados para análise ou proposição de tesouros interoperáveis semanticamente. A importância de elaborar tesouros interoperáveis fica mais evidente diante da forma como a Web tem impactado o uso de tesouros, visto que “com a heterogeneidade de formatos e estruturas na web, os tesouros precisam incorporar modelos de dados flexíveis e que possam ser estendidos para melhor representar seu conteúdo” (Barbosa; Viera, 2022, p. 232).

Com efeito, a construção de tesouros precisa passar por uma mudança de paradigma, em que a criação desses instrumentos terminológicos não seja focada apenas em termos, mas com enfoque maior em conceitos, promovendo assim, a operacionalização de tesouros conceituais em que os conceitos serão fundamentais para uso combinado dos tesouros internamente e com outros vocabulários e sistemas interoperáveis (Mathews; Miles; Wilson, 2009).

Os sistemas convencionais de organização do conhecimento, que incluem tesouros, esquemas de classificação, taxonomias, sistemas de cabeçalho de assunto, autoridades de nomes e listas de códigos e termos, têm demonstrado uma rápida adaptação às novas tecnologias em diferentes contextos de publicação e aplicação. Zeng e Mayr (2018) apontam que desde a década de 1950, eles têm sido utilizados em bancos de dados de indexação e resumos, serviços de informação *online*, CD-ROMs, arquivos Adobe PDF, sites da Web HTML e bancos de dados XML. Mais recentemente, os SOC estão sendo estruturados para sua integração à Web Semântica, contribuindo para o desenvolvimento de produtos em conjuntos de dados LOD, juntamente com ontologias desenvolvidas no século XXI (Zeng; Mayr, 2018).

Zeng e Mayr (2018) defendem que ao criar representações das estruturas semânticas na Web que fundamentam os domínios, os SOC desempenham o papel de mapas semânticos, fornecendo uma referência compartilhada para indexadores e usuários, sejam eles humanos ou máquinas, permitindo uma orientação comum.

---

<sup>7</sup> *Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval*

<sup>8</sup> *Information and documentation — Thesauri and interoperability with other vocabularies — Part 2: Interoperability with other vocabularies*

Portanto, o tesouro, no contexto da ORC atrelada à Ciência da Informação, é compreendido simultaneamente como um instrumento terminológico, um SOC e uma linguagem documentária.

#### **4 SIMPLE KNOWLEDGE ORGANIZATION SYSTEM E NORMA ISO 25964**

Compreender o SKOS a partir de sua documentação, autores nacionais e internacionais é fundamental para captar conceitos que condicionam a análise de abordagens e aplicações de tesouros representados em SKOS. Deste modo, nesta seção é dada continuidade à apresentação do referencial teórico mais restrito ao modelo de dados SKOS e o contexto em que está inserido, atentando-se para similaridades entre a estrutura SKOS e a estrutura de tesouros clássicos. Nesta perspectiva, a subseção 4.1 apresenta o contexto em que o referido modelo foi desenvolvido e respectivas características estruturais. Por sua vez, a subseção 4.2 apresenta os aspectos normativos com relação à construção de tesouros, no tangente à norma ISO 25964-1:2011 (tesauro para recuperação da informação) e ISO 25964-2:2013 (tesauro e interoperabilidade com outros vocabulários), de modo a embasar as análises das similaridades entre as recomendações do modelo de dados SKOS e as recomendações da referida norma, apresentadas e discutidas na Seção 5.

##### **4.1 Simple Knowledge Organization System**

Para Baeza-Yates e Ribeiro-Neto (2013) os desafios de recuperação da informação no contexto da Web podem ser divididos em duas frentes: desafios de recuperação centrados em dados e desafios da recuperação centrados em usuários. Manning, Raghavan e Schutze (2008) pontuam que a Web não possui precedentes em sua escala e na quase total falta de coordenação em sua criação. O crescimento exponencial do seu uso abarca uma grande parcela da humanidade ao apostar em um *design* simples de cliente-servidor. De maneira geral, esta relação de cliente-servidor ocorre quando um servidor da Web se comunica com um usuário por meio de um protocolo de transferência de hipertexto seguro (HTTPS<sup>9</sup>). Este processo de transferência ocorre de maneira leve e simples e é capaz de transportar uma diversidade de dados como texto, imagens e arquivos de áudio e vídeo. Porém, a grande massa de informação disponível na Web não possui utilidade se ela não for descoberta e consumida por outros usuários (Manning; Raghavan; Schutze, 2008). A existência dessa massa de informação, já levou ao cunho do termo “web de

---

<sup>9</sup> *Hyper Text Transfer Protocol Secure*

dados” (Catarino, 2014), que é diferente do ambiente web que em seus primórdios não admitia a produção e consumo exponencial de dados e documentos de forma tão facilitada como atualmente.

Este cenário relembra a Documentação, em que Suzanne Briet, principal discípula de Paul Otlet, apresenta documento como sendo qualquer tipo de recurso que informe a partir de evidências (Smit, 2008). No entanto, como atualmente há um conjunto gigantesco e crescente de dados na Web, fica difícil definir o que realmente interessa enquanto recurso informacional de acordo com necessidades e contextos específicos. Esta dificuldade também remonta à segunda e à terceira lei da Biblioteconomia: “para cada leitor o seu livro” e “para cada livro o seu leitor”. Ao fazer uma adaptação para o contexto do imenso e heterogêneo volume de dados disponíveis na Web, tem-se: *para cada usuário o seu conjunto de dados* e *para cada conjunto de dados o seu usuário*. Seguindo esta ideia para a primeira, quarta e quinta lei, respectivamente, *os dados são para serem usados, poupe o tempo do usuário e a Web é um organismo em crescimento*. Sendo que esta última relembra a corrente epistemológica cibernética de segunda ordem apontada por Capurro (2003). Pelas lembranças trazidas neste parágrafo, percebe-se o quanto as origens da Ciência da Informação estão presentes e são necessárias para analisar e propor soluções para os desafios próprios do ambiente digital na Web.

A Web<sup>10</sup> é uma rede que foi inventada por Tim Berners-Lee e Robert Cailliau em 1989. Tecnicamente e por meio do uso da Internet, essa rede surge com o objetivo de permitir a consulta e atualização de documentos disponibilizados em uma estrutura de hipertexto, em que suas páginas são estruturadas a partir de *links* que interligam diferentes recursos informacionais, por exemplo, textos, imagens, vídeos, dentre outros. Arquitetada com base no conceito de cliente servidor, em que uma aplicação (um cliente Web) requisita um documento (um recurso informacional) a uma outra aplicação (um servidor Web) que informa e identifica esse documento (Laufer, 2015).

Constatada a quantidade imensa de conteúdos criados e compartilhados nesse ambiente, seu crescimento exponencial e com o propósito de estender a Web para ambientes semânticos, onde os dados poderão ser legíveis e processáveis

---

<sup>10</sup> <https://www.w3.org/History/1989/proposal.html>

semanticamente por algoritmos de busca, o W3C tem colaborado com a criação de um conjunto de tecnologias para permitir que as máquinas de busca compreendam adequadamente os dados atrelados aos recursos informacionais, bem como o significado das relações entre dois ou mais recursos (Isotani; Bittencourt, 2015). Diante desse intento de compreensão e significado por entre as máquinas e recursos de informação, tem-se a Web Semântica como um projeto de ambientes semânticos no cerne da Web e os heterogêneos conjuntos de dados através dela disponibilizados, onde a estruturação de dados linkados (*Linked Data*) é condição para tornar os dados da Web legíveis por humanos e máquinas (Baeza-Yates; Ribeiro-Neto, 2013; Catarino, 2014; Laufer, 2015). Neste sentido, Clarke e Zeng (2012, p. 20) colocam que “para atuar na Web Semântica, o software de computador precisa de um modelo de dados explícito que distinga entre termos e conceitos”.

A Web Semântica<sup>11</sup>, enquanto conjunto de ambientes semânticos, provê uma estrutura semântica para páginas Web. O termo Web Semântica refere-se à visão do W3C de Web de *Linked Data*, reconhecido como um movimento que visa um conjunto de práticas “que propõe a ligação de dados representados em um formato padrão, a partir da utilização de links semânticos, da utilização de ferramentas e de práticas propostas pelo movimento” (Castro; Jesus, 2018, p. 150). Esse movimento se expande como LOD (Pastor-Sánchez, 2013b; Zeng, 2018), focado no acesso aberto aos dados (Laufer, 2015; Zeng, 2018), regido pelos seguintes princípios: as entidades devem ser nomeadas por meio de

[...] URIs únicas; essas URIs devem ser URIs HTTP e serem resolvidas usando protocolos web padrão; quando essas URIs são resolvidas, devem retornar informações úteis sobre o recurso; elas devem conter links para outras URIs para que as pessoas possam descobrir recursos relacionados (Martínez-González; Alvite-Díez, 2019b, p. [6], tradução livre).

Deste modo, a Web Semântica, através de suas tecnologias e padrões semânticos, define a forma como sintaticamente os metadados podem ser agregados às informações por meio do uso de vocabulários que possuam uma semântica bem definida para que o significado pretendido pelo publicador seja o mesmo que o significado entendido pelo consumidor dos dados.

---

<sup>11</sup> <https://www.w3.org/standards/semanticweb/>

Durante as décadas de 1990 e 2000, com o objetivo de promover o desenvolvimento, aplicação e interação de tesouros na Web Semântica, o W3C, por meio do grupo de trabalho *SWAD-Europe Thesaurus Activity*, se direciona para o desenvolvimento e aplicação de tecnologias para este fim. Entre essas tecnologias estão: a linguagem *Extensible Markup Language* (XML), a linguagem *Ontology Web Language* (OWL) e o modelo de dados *Resource Description Framework*<sup>12</sup> (RDF) (Barbosa; Viera, 2022). Nesse cenário, o grupo de trabalho *Semantic Web Advanced Development for Europe* (SWAD-E) iniciou o desenvolvimento do modelo *Simple Knowledge Organization System* (SKOS), com o objetivo de que os KOS tradicionais sejam registrados e legíveis por algoritmos (Isaac; Summers; 2009). Em 2005, o SKOS foi divulgado a partir da publicação do *SKOS Core Guide* (Pastor-Sanchez; Martínez-Mendez; Rodriguez-Muños, 2009).

Tempo depois, no ano de 2009 o W3C formalizou oficialmente a Recomendação do padrão SKOS como um modelo de dados para a representação de SOC no ambiente Web (Miles, Bechhofer, 2009; Pastor-Sanchez; Martínez-Mendez; Rodriguez-Muños, 2009; Ramalho, 2015; Zeng, 2018). Conforme Catarino, Cervantes e Andrade (2015, p. 110), o “*Simple Knowledge Organization System* (SKOS) é um modelo para expressar a estrutura básica e conteúdo de sistemas de organização do conhecimento” como os vocabulários controlados e tesouros, constituindo-se em um sistema simples para a organização do conhecimento, em que as estruturas baseadas em SKOS se manifestam de forma simples, pois não produzem inferências lógicas e formais como as ontologias.

Deste modo, a integração entre a iniciativa ou movimento LOD e a implementação de soluções interoperáveis é válida, ao passo que

A permeabilidade dos vocabulários na web contribui para a evolução de uma web de dados. Uma vez modelados e interligados, os vocabulários ampliam não apenas a capacidade de troca semântica entre sistemas de computador, como também, permitem que os usuários ampliem suas possibilidades de busca e de recuperação da informação (Barbosa; Viera, 2022, p. 231).

Zeng e Mayr (2018, p. 2, tradução livre) se utilizam da expressão “LOD KOS” para designar “vocabulários de valor e ontologias leves dentro do framework da Web

---

<sup>12</sup> <https://www.w3.org/TR/rdf-primer/>

Semântica”. Os autores apontam que esse entendimento terminológico advém das comunidades LOD, para as quais os sistemas de organização do conhecimento são usados como “vocabulários de valor”, pois diferentemente dos “vocabulários de propriedade” (conjuntos de elementos de metadados), referem-se ao seus usos “nos modelos baseados em RDF, nos quais os triplos "recurso, tipo de propriedade, valor da propriedade" se beneficiam de uma lista controlada de valores permitidos para um elemento em dados estruturados” (Zeng; Mayr, 2018, p. 2, tradução livre).

Um SOC baseado em LOD deve aderir aos princípios de *Linked Data* e estar amplamente disponível. O modelo de dados SKOS define um SOC como um esquema conceitual composto por um conjunto de conceitos, sendo que cada conceito deve ter um nome representado por um *Identificador de Recurso Uniforme* (URI) ou Identificador de Recurso Internacionalizado (IRI) (Zeng; Mayr, 2018). O URI é uma forma de identificar semanticamente cada um dos recursos e respectivas propriedades de modo único e universal (Laufer, 2015), colocando-se como um elemento importante para os processos de ORC.

A utilização de identificadores únicos para representar entidades ou recursos é uma solução fundamental para fornecer dados processáveis por máquina e sem ambiguidades. Além disso, ao disponibilizar um conjunto de dados como LOD, é recomendado o uso de URIs HTTP. Os dados de um “LOD KOS” são expressos como triplas RDF e podem ser codificados em diferentes sintaxes (serializações) RDF, como RDF/XML, Turtle, TriG, N-Quads e JSON-LD. Isso permite que os dados sejam transmitidos de forma interoperável entre aplicativos de computador, facilitando o uso de um SOC em aplicativos de metadados distribuídos e descentralizados (Laufer, 2015; Zeng; Mayr, 2018).

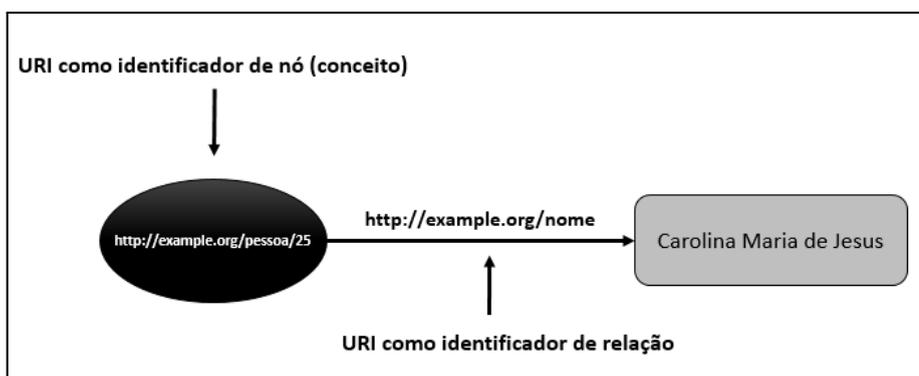
A apresentação formal de um SOC do tipo tesouro que se configure como um “LOD KOS” é uma das principais viabilidades proporcionadas pelo modelo SKOS, que tem como principais atributos: estrutura legível por máquina; codificação no padrão RDF – com codificação estruturada em recursos (*Subjects*), propriedades (*Predicate*) e valores (*Object*); alternativa de baixo custo que permite a transição de Sistemas de Organização do Conhecimento tradicionais para o ambiente Web; provê interoperabilidade entre SOC; condicionamento para recuperação semântica de informações na Web; permite identificação de conceitos a partir do padrão *Uniform*

*Resource Identifiers* (URI) e; estrutura categorizada em conceitos, propriedades e relações (Catarino 2014; Miles; Bechhofer, 2009; Ramalho, 2015; Zeng; Mayr, 2018).

A codificação SKOS é baseada no modelo de dados RDF e pode ser aplicada em linguagem XML, enquanto “sintaxe para o intercâmbio e o processamento de metadados” (Ferreira; Santos, 2013, p. 13). O RDF é um arcabouço que se constitui numa linguagem que estrutura a representação de recursos na Web Semântica. Os *links* descritos nessa linguagem são utilizados para acessar (linkar) dados de diversos recursos. O RDF permite fazer afirmações sobre quaisquer recursos, tanto concretos quanto abstratos. Uma determinada empresa, uma pessoa, uma página Web, um sentimento, uma cor, também são considerados recursos (Laufer, 2015).

A codificação RDF descreve a relação entre recursos a partir de uma estrutura de triplas do tipo <sujeito> <predicado> <objeto> - onde sujeito = *Subjects* = recurso; propriedades = *Predicate* = predicado; e valores = *Object* = objeto (Isotani; Bittencourt, 2015). O conjunto destas estruturas em triplas é chamado de Grafo RDF. A representação de um grafo RDF pode ser visualizada na Figura 3, a saber:

**Figura 3 - Tripla RDF.**



Fonte:Elaborado pela autora.

Para que uma URI não se perca com o tempo é fundamental que ela seja elaborada de forma persistente (Zeng; Mayr, 2018).

Relembrando preceitos da gramática, como numa oração coordenada, a Figura 3 expressa uma relação entre dois recursos - *http://example.org/person/25* como sujeito e *Carolina Maria de Jesus* como objeto (valor), daí a designação “vocabulário de valor” para denominar SOC baseados em modelo de dados RDF

(Zeng; Mayr, 2018). O predicado nome (*http://example.org/nome*) representa a natureza desta relação, que é formulada de modo direcional (do sujeito para o objeto). Em RDF, essa representação é chamada de propriedade (Laufer, 2015). Assim, as relações entre sujeito e objeto são indicadas através de predicados.

O modelo de dados SKOS, então, é uma aplicação do modelo RDF e, a partir do entendimento deste, é possível entender detalhadamente os elementos básicos que compõem o SKOS: conceitos, propriedades e relações. Nesse sentido, no SKOS, os conceitos fazem o papel dos recursos (sujeitos), enquanto as propriedades (atributos) são os predicados que expressam relações entre conceitos e respectivos valores ou entre dois conceitos por meio de relações hierárquicas, associativas e de equivalência. Assim, é válido ressaltar que um recurso pode estabelecer uma relação com outro recurso e não apenas com o valor literal de uma propriedade.

#### **4.1.1 Conceitos no modelo SKOS**

Quando um tesouro está representado em SKOS, as relações entre seus conceitos (termo/sujeito) estarão modeladas em RDF, o que tornará as relações hierárquicas, de equivalência e associativas passíveis de serem lidas por agentes inteligentes (Isaac, Summers, 2009; Catarino, 2014; Miles, Bechhofer, 2009). Nesse sentido, o padrão URI atrelado ao modelo SKOS permite que os conceitos sejam representados (Catarino, 2014; Ramalho, 2015; Zeng, 2018), enquanto o RDF comporta a modelagem da declaração de predicados e valores dos conceitos. As triplas RDF codificadas em linguagem artificial com sintaxe específica (um código como, por exemplo, a linguagem Turtle<sup>13</sup>) permitem que o conceito enquanto termo (valor) do vocabulário em SKOS possa ser processado automaticamente pelos agentes inteligentes computacionais (algoritmos de busca, por exemplo) de acordo com as relações terminológicas/conceituais estabelecidas no âmbito do vocabulário de um determinado domínio do conhecimento (Isaac, Summers; Miles, Bechhofer, 2009). Os conceitos são elementos centrais dos SOC, inclusive quando estes estão estruturados em SKOS. Para tanto, em SKOS há “uma classe que define que um determinado recurso é um conceito” (Laufer, 2015, p. 96), por exemplo:

---

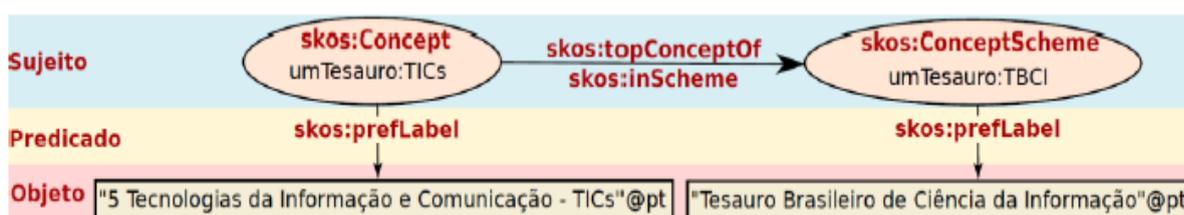
<sup>13</sup> O código Turtle é uma forma de expressar as triplas RDF. Ele permite representar um grafo RDF em uma forma compacta baseada em texto, tornando mais fácil a edição dos dados. O código Turtle pode ser colocado em um arquivo com extensão ".ttl".

@prefix ex: <http://example.com>.

ex:books rdf:type skos:Concept.

A notação *skos:Concept* é a que designa a classe de conceitos no âmbito do SKOS, enquanto a classe *skos:ConceptScheme* agrega um conjunto de conceitos (*skos:Concept*) de um domínio do conhecimento. Essa agregação aponta que uma instância da classe *skos:ConceptScheme* pode ser um SOC como, por exemplo, o próprio tesauro ou um microtesauro. Segue uma ilustração de relacionamento entre uma instância da classe *skos:Concept* e uma instância da classe *skos:ConceptScheme*:

**Figura 4** - Relacionamento entre *skos:Concept* e *skos:ConceptScheme*.



Fonte: Rautenberg, Souza e Kelnar (2020).

Na Figura 4 tem-se a ilustração de quatro triplas RDF (<sujeito> <predicado> <objeto>): a) <umTesauro:TICS> <skos:prefLabel> <"5 Tecnologias da Informação e Comunicação - TICS"@pt>; b) <umTesauro:TBCI> <skos:prefLabel> <"Tesauro Brasileiro de Ciência da Informação"@pt>; c) <umTesauro:TICS> <skos:topConceptOf> <umTesauro:TBCI>; e d) <umTesauro:TICS> <skos:inScheme><umTesauro:TBCI>.

Com relação às notações SKOS para designar, relacionar e definir conceitos e esquemas de conceitos, segue o exposto no Quadro 4:

**Quadro 4** - Funções das notações para conceitos e esquemas de conceitos.

Notação	Função
<i>skos:ConceptScheme</i>	agregar um conjunto de <i>skos:Concept</i> de um domínio particular.
<i>skos:Concept</i>	representar as unidades de conhecimento (uma ideia, conceito a ser valorado por um termo) de um domínio.

<i>skos:hasTopConcept</i>	mapear os <i>skos:Concept</i> mais genéricos (ou importantes), considerando um <i>skos:ConceptScheme</i> .
<i>skos:topConceptOf</i>	relacionar inversamente o <i>skos:hasTopConcept</i> que mapeia um <i>skos:Concept</i> mais genérico em um <i>skos:ConceptScheme</i> .
<i>skos:inScheme</i>	identificar quais <i>skos:Concept</i> pertencem a um <i>skos:ConceptScheme</i> .
<i>skos:definition</i>	descrever um <i>skos:Concept</i> . Pode-se apresentar várias definições de acordo com os idiomas de referência (@pt, @en, @es, dentre outras).

Fonte: adaptado de Rautenberg, Souza e Kelniar (2020).

De acordo com o Quadro 4, a notação *skos:topConceptOf* demonstrada na Figura 4 indica que a etiqueta (literal/ termo) “5 Tecnologias da Informação e Comunicação - TICs”@pt, para o sujeito umTesauro:TICs, aponta um esquema de conceitos inserido num esquema de conceitos maior (representado pelo literal “Tesauro Brasileiro de Ciência da Informação”@pt). Nesse sentido, o sujeito umTesauro:TICs pode ser considerado um microtesauro.

#### 4.1.2 Propriedades no modelo SKOS

As propriedades SKOS descrevem os atributos dos conceitos do vocabulário. Existem propriedades de dois tipos: propriedades de etiquetagem e propriedades de documentação. As propriedades do primeiro tipo, em notação padronizada pelo SKOS, são *skos:prefLabel*, *skos:altLabel* e *skos:hiddenLabel* (Catarino 2014; Miles; Bechhofer, 2009; Ramalho, 2015). Estas propriedades, conforme esclarecimentos anteriores sobre o modelo RDF, estabelecem relações entre os conceitos (enquanto recursos abstratos ou concretos) e valores que designam uma etiqueta (termo) para o conceito. Ao fazer uma analogia com as notações dos tesouros, essas propriedades de etiquetagem podem ser compreendidas como propriedades que se constituem em notações terminológicas para os termos enquanto representantes de conceitos dos tesouros. As propriedades de etiquetagem *skos:prefLabel* e *skos:altLabel* podem ser comparadas respectivamente com a designação para “termo preferido” (indicado pela notação tradicional USE) e o “termo alternativo” (“termo não preferido” indicado pela notação tradicional UP). A notação *skos:hiddenLabel*, mesmo não possuindo uma notação análoga em tesouros

clássicos, também indica um tipo de notação terminológica por designar “termo oculto”. Logo, as propriedades de etiquetagem do SKOS possibilitam representar relações semânticas a nível lexical, relação entre conceito e termo. As etiquetas (literais/ termos) SKOS podem ser declaradas em diferentes idiomas.

Segue descrição das propriedades de etiquetagem no Quadro 5:

**Quadro 5** - Notações e propriedades de etiquetagem do modelo SKOS.

Notação	Propriedade de Etiquetagem	Descrição
<i>skos:prefLabel</i>	Etiqueta Preferencial	Define o termo preferencial de um conceito em um determinado idioma.
<i>skos:altLabel</i>	Etiqueta Alternativa	Define termos alternativos de um conceito, como sinônimos, acrônimos, abreviações, variações de ortografia, e formas de plural/singular.
<i>skos:hiddenLabel</i>	Etiqueta Oculta	Define termos referentes a um conceito que devem ser ocultos na geração de apresentações visuais, mas que podemos facilitar o acesso ao conceito em operações de busca livre. Erros de ortografia e digitação são normalmente incluídos a partir desta etiqueta.

Fonte: adaptado de Ramalho (2015).

Em atenção às descrições do Quadro 5, observa-se que as etiquetas permitem a expressão dos conceitos em linguagem natural (Laufer, 2015) como, por exemplo:

```
ex:books rdf:type skos:Concept;
    skos:prefLabel "books";
    skos:altLabel "publications"
```

As propriedades de documentação são responsáveis por documentar as características mais gerenciais dos conceitos do vocabulário em SKOS, de modo que sejam armazenadas informações importantes para a gestão do vocabulário. Dentre as propriedades de documentação está a nota histórica, representada pela notação *skos:historyNote*, servindo para descrever o histórico de mudanças significativas do significado, ou formas de utilização de um conceito, por exemplo (Catarino, 2014; Isaac, Summers; Miles, Bechhofer, 2009; Ramalho, 2015).

### 4.1.3 Relações semânticas entre conceitos no modelo SKOS

Durante análise das recomendações para o modelo SKOS, a partir da literatura científica e documentos do W3C (Isaac; Summers, 2009; Miles; Bechhofer, 2009), percebeu-se que as relações semânticas principais ocorrem entre os conceitos (recursos conceituais) e não entre conceitos e valores (*strings* em linguagem natural) de um esquema de conceitos identificados por URIs. Porém, todas as relações indicadas a partir das notações SKOS contribuem para expressar a semântica estrutural de um tesauro em SKOS.

Para Catarino (2014) e Ramalho (2015), essas relações semânticas entre conceitos são do tipo relações hierárquicas, relações associativas ou relações de equivalência, assim como as existentes nos tesouros, porém, nesse contexto em que são modeladas em SKOS, tornam-se legíveis por agentes inteligentes. Exemplos das notações dessas relações estão descritas no Quadro 6, em que as relações de equivalência são as relações que mais contribuem para a interoperabilidade entre diferentes SOC construídos sob o modelo SKOS (Isaac, Summers, 2009; Ramalho, 2015).

**Quadro 6** - Notações de relações semânticas entre conceitos no modelo SKOS.

Tipo de relação	Notação	Descrição
Relações Hierárquicas	<i>skos:broader</i>	Define Relações Hierárquicas entre conceitos, indicando que determinado conceito possui um significado mais amplo. Equivalente à notação TG no tesauro clássico (Termo Genérico).
	<i>skos:narrower</i>	Define Relações Hierárquicas entre conceitos, indicando que determinado conceito possui um significado mais específico. Equivalente à notação TE no tesauro clássico (Termo Específico).
Relações Associativas	<i>skos:related</i>	Define Relações Associativas entre conceitos. Equivalente à notação TR no tesauro clássico (Termo Relacionado).

Relações de Equivalência	<i>skos:exactMatch</i>	Define Relações de Equivalências entre conceitos que possuem alto grau de correspondência e podem ser utilizados indistintamente em uma ampla gama de aplicações de recuperação da informação. Este tipo de relação relembra as relações paradigmáticas das LDs de um modo geral.
	<i>skos:closeMatch</i>	Define Relações de Equivalências entre conceitos que podem ser considerados como similares em contexto previamente determinado. Este tipo de relação relembra as relações sintagmáticas das LDs de um modo geral.
	<i>skos:broadMatch</i>	Define Relações de Equivalências, considerando a estrutura hierárquica de um conceito que possui um significado mais amplo.
	<i>skos:narrowMatch</i>	Define Relações de Equivalências, considerando a estrutura hierárquica de um conceito que possui um significado mais específico.
	<i>skos:relatedMatch</i>	Define Relações de Equivalências, considerando as estruturas associativas existentes entre conceitos.

Fonte: adaptado de Ramalho (2015) e Rautenberg, Souza e Kelnar (2020).

Diante dos esclarecimentos, o modelo SKOS compreende um padrão para apoiar a recuperação semântica da informação. O modelo de dados SKOS é o primeiro passo para alcançar a interoperabilidade semântica entre os vocabulários e tornou-se um padrão reconhecido para a publicação de tesouros em ambientes semânticos da Web (Zeng; Mayr, 2018). Este modelo permite a representação das estruturas de diversos tipos de SOC, oferecendo, assim, a possibilidade de mapear conceitos para outros vocabulários, por isso tem sido amplamente utilizado como formato de representação de vocabulários na Web (Bezares; Rodríguez; Fuente, 2020).

#### 4.1.4 Exemplos de tesouros representados em SKOS

A literatura científica destaca alguns tesouros representados em SKOS, a citar: Tesouro da Unesco, EuroVoc e AGROVOC, conforme apresentado a seguir:

- AGROVOC<sup>14</sup>

Na década de 1980, a Organização das Nações Unidas por meio da *Food and Agriculture Organization* (FAO) publicou o AGROVOC pela primeira vez em inglês, espanhol e francês. Essa publicação teve como objetivo principal servir como uma ferramenta de indexação para a biblioteca da FAO e o banco de dados *International Information System for the Agricultural Sciences and Technology* (AGRIS), que é o Sistema Internacional de Ciência e Tecnologia Agrícola. O AGROVOC é um tesouro abrangente que aborda todas as áreas de interesse da FAO (alimentos, nutrição, agricultura, pesca, silvicultura, meio ambiente, entre outros). Atualmente, o AGROVOC é mantido por uma comunidade ampla de especialistas e instituições, com o suporte e a coordenação técnica da FAO. Regularmente, são lançadas atualizações do conteúdo, sendo a versão mais recente a 3.4. O AGROVOC é amplamente utilizado em bibliotecas especializadas, bem como em bibliotecas digitais e repositórios, com o propósito de indexar conteúdo e realizar mineração de texto. No ano 2000, o AGROVOC abandonou a impressão em papel e passou a ser disponibilizado em formato digital. Em 2009, ele se tornou um recurso SKOS. Hoje, o AGROVOC está disponível como um esquema de conceitos SKOS-XL e publicado como um conjunto de dados LOD, contendo mais de 36.000 conceitos disponíveis em até 33 idiomas. A edição do AGROVOC é realizada por meio do uso do *software* VocBench. O tesouro AGROVOC possui uma organização hierárquica composta por 25 conceitos principais. Esses conceitos principais são de natureza geral e abrangente, incluindo termos como "atividades", "entidades", "locais", "produtos", "substâncias", "organismos", entre outros. O fato de mais de 20.000 conceitos se enquadrarem no conceito principal "organismos" confirma o foco do AGROVOC no setor agrícola. Além disso, o AGROVOC está alinhado com outros 18 SOC multilíngues e pode ser acessado por meio de um ponto de acesso (ou terminal) *SPARQL Protocol and RDF Query Language* (SPARQL). Existem duas opções de *download* disponíveis: CORE, que inclui apenas os conceitos do

---

<sup>14</sup> <https://agrovoc.fao.org/browse/agrovoc/en/>

AGROVOC em formato RDF e NT, e LOD, que inclui os conceitos do AGROVOC juntamente com *links* para recursos externos em formato NT e NQ. A publicação do AGROVOC na web utiliza o Skosmos como ferramenta de busca e navegação. Além disso, há uma ontologia associada a esse tesouro, chamada Agrontologia (Alvite-Díez; Martínez-González, 2021; Berti Junior *et al.*, 2017; Caracciolo *et al.*, 2012; Martínez-González; Alvite-Díez, 2019b,c).

- EuroVoc<sup>15</sup>

Tesouro multilíngue desenvolvido originalmente para processar informações documentais das instituições da União Europeia (UE). No ano de 1982, o Parlamento Europeu e a Comissão Europeia optaram por construir um tesouro multilíngue abrangendo os campos de interesse das instituições europeias e que estivesse em conformidade com os padrões internacionais pertinentes. A primeira edição do tesouro EuroVoc foi publicada em 1984 em sete idiomas. Posteriormente, após edições impressas consecutivas, a quarta edição foi disponibilizada na Internet em 2002. Desde a Edição 4.4, o EuroVoc tem sido publicado e disponibilizado na Internet em 23 idiomas da UE. O tesouro EuroVoc é continuamente adaptado e sua edição, assim como o tesouro AGROVOC, também é realizada por meio do uso do VocBench. A versão atual, 4.9, foi publicada em dezembro de 2018. O EuroVoc é um tesouro multidisciplinar que abrange campos suficientemente amplos para incluir perspectivas tanto da comunidade de tesouros quanto perspectivas nacionais, com ênfase nas atividades parlamentares. O EuroVoc é gerenciado pelo Escritório de Publicações da UE e seus usuários incluem as instituições da UE, o Escritório de Publicações da UE, parlamentos nacionais e regionais na Europa, além de governos nacionais e usuários privados em todo o mundo. O tesouro EuroVoc foi compilado em conformidade com o padrão ISO 25964. Ele é dividido em 21 domínios e 127 microtesouros. Cada domínio é subdividido em vários microtesouros. A versão atual contém 7.222 conceitos. O EuroVoc está alinhado com outros 15 sistemas de organização do conhecimento e pode ser baixado em formatos XML, SKOS-Core, SKOS-AP-EU (formato RDF desenvolvido pelo Escritório de Publicações), Excel e Marc-XML. Além disso, o EuroVoc também possui uma ontologia chamada EuVoc1,

---

<sup>15</sup> <https://eur-lex.europa.eu/browse/eurovoc.html?locale=pt>

que estende o SKOS com classes e propriedades específicas para esse tesauro (Alvite-Díez; Martínez-González, 2021; Martínez-González; Alvite-Díez, 2019b,c).

- Tesauro da UNESCO<sup>16</sup>

Foi elaborada uma versão preliminar do Tesauro da *United Nations Educational, Scientific and Cultural Organization* (UNESCO) entre 1973 e 1974, sendo que a primeira edição em inglês foi lançada em 1977, seguida das traduções para o francês e espanhol em 1983 e 1984, respectivamente. Uma segunda versão revisada e reestruturada foi disponibilizada em 1995. Atualmente, o tesauro está disponível em inglês, francês, russo e espanhol. O tesauro da UNESCO é amplamente utilizado para indexar e buscar recursos no repositório da UNESCO, relacionados a documentos e publicações nas áreas de educação, cultura, ciências naturais, ciências sociais e humanas, comunicação e informação. Trata-se de um vocabulário multidisciplinar que é constantemente enriquecido e atualizado. O Tesauro da UNESCO também está em conformidade com o padrão ISO 25964. Ele é composto por 4.417 conceitos, agrupados em 7 domínios temáticos ("Educação", "Ciência", "Cultura", "Ciências Sociais e Humanas", "Informação e Comunicação", "Política, Direito e Economia" e "Países e Agrupamentos de Países"), além de possuir 88 microtesauros. O tesauro foi publicado como um conjunto de dados SKOS, seguindo os princípios de LOD, e está disponível para *download* nas versões SKOS Core e SKOS-XL. Os formatos de *download* oferecidos incluem PDF, RDF/XML, TURTLE e JSON-LD. Além disso, o acesso ao Tesauro da UNESCO é possível por meio de um ponto de acesso SPARQL (*SPARQL Protocol and RDF Query Language*). A versão oficial desse tesauro utiliza a ontologia ISO-THES (Alvite-Díez; Martínez-González, 2021; Martínez-González; Alvite-Díez, 2019b,c).

Além destes três tesouros destacados, existem outros tesouros e esquemas conceituais cujas representações se baseiam no modelo SKOS e na norma ISO 25964. Dentre estes, o tesauro Zonom (Saliba *et al.*; 2021), apresentado na subseção 5.12.4, pois integra parte dos resultados da presente pesquisa.

“Os tesouros são os vocabulários controlados com maior adesão ao modelo SKOS” (Barbosa; Viera, 2022, p. 231). Apesar disso, existem algumas

---

<sup>16</sup> <https://vocabularies.unesco.org/browser/thesaurus/en/>

especificidades quanto à aplicação do modelo SKOS para a representação de SOC do tipo tesouros. Essas especificidades estão relacionadas ao refinamento dos relacionamentos semânticos, estruturação hierárquica dos conceitos em vocabulários multidisciplinares e interoperabilidade com outros conjuntos de dados ou esquemas de conceitos. Por isso, surgiram modelos de extensão do SKOS *Core* original com o propósito de colaborar com o aprimoramento das relações semânticas e com a interoperabilidade intra e inter-tesouros na Web (Barbosa; Viera, 2022). Esses modelos de extensão também visam atender à norma ISO 25964 (que é específica para tesouros) e, por vezes, se constituem como soluções *ad hoc* para tesouros e outros tipos SOC desenvolvidos em ambientes semânticos da Web. Conforme os modelos de tesouros apresentados acima, fora citada a extensão SKOS-XL (*SKOS eXtension for Labels*), mas também há a *UNESKOS Vocabulary Specification* e a ISO-THES, que estão entre as tecnologias para aprimoramento do modelo SKOS. O primeiro surge como modelo de extensão para mitigar as limitações originais do SKOS para representações de SOC do tipo tesouros. Enquanto o segundo e o terceiro surgem como modelos de extensão para complementar o SKOS de acordo com a norma ISO 25964 (Pastor-Sánchez, 2016; Barbosa; Viera, 2022).

#### **4.2 Aspectos normativos: norma ISO 25964**

Uma norma se constitui num padrão, pois regula procedimentos, estabelece regras e princípios que direcionam e refletem técnicas e costumes acordados por uma comunidade. Para Martínez-González e Alvite-Díez (2019c, p. 79, tradução livre) “normas são os instrumentos que coletam o acordo de uma comunidade sobre a questão padronizada”. Em relação aos tesouros, a norma ISO 25964 (parte 1:2011, tesouro para recuperação da informação, e parte 2:2013, tesouro e interoperabilidade com outros vocabulários) “define o que é um tesouro e quais são as construções que um tesouro pode usar para ser um vocabulário controlado válido” (Martínez-González; Alvite-Díez, 2019c, p. 79, tradução livre) para estruturação e representação de dados semânticos, interoperáveis e legíveis por algoritmos.

Por um longo período, as diretrizes ISO 5964:1985 e ISO 2788:1986 foram as únicas fontes normativas relevantes para o desenvolvimento e manutenção de

tesauros. Embora tenha havido algumas revisões parciais em certos momentos, as alterações implementadas não trouxeram mudanças significativas à estrutura e aos componentes do tesauro (Bandyopadhyay; Mukhopadhyay, 2015; Clarke; Zeng, 2012; ISO TC46/SC9/WG8; Isaac, 2012; Pastor-Sánchez, 2013a). Porém surgem as cinco partes da norma britânica BS 8723 (2005-2008) e, posteriormente, as duas partes da norma ISO 25964 (2011 e 2013) que trazem mudanças substanciais para a composição de tesauros (Clarke; Zeng, 2012; Pastor-Sánchez, 2013b). Em especial, a norma ISO 25964 (partes 1 e 2) se aproxima mais do modelo de dados SKOS, pois diferentemente das normas anteriores, que focam no controle léxico de vocabulário, a norma vigente visa os níveis de representação conceitual do SOC do tipo tesauros. Clarke e Zeng (2012) apontam que a necessidade de elaboração da norma não foi totalmente reconhecida até o final do século XX. Até este período, os tesauros eram amplamente empregados em contextos nos quais a busca de informações era controlada ou mediada por seres humanos. Nesses casos, os usuários humanos tinham a capacidade intuitiva de distinguir entre um termo e um conceito (parte do processo de semiose), e conseguiam interpretar os resultados da pesquisa sem confusão. A necessidade de um modelo de dados surge somente quando uma máquina precisa de instruções claras sobre como manipular e interpretar os dados (Clarke; Zeng, 2012).

#### Quadro 7 - Silogismo com termo e conceito.

<p>Homem é uma palavra de 5 letras. (<i>homem</i> como termo)</p> <p>Sócrates é um homem. (<i>homem</i> como conceito)</p> <p>Portanto, Sócrates é uma palavra de 5 letras. (inferência lógica errônea)</p>	<p>A falha lógica é muito óbvia para um leitor humano, mas um computador pode ser facilmente enganado se declarações sobre um termo forem apresentadas como declarações sobre o conceito representado pelo termo.</p>
---	---

Fonte: adaptado de Clarke e Zeng (2012).

A partir do Quadro 7, percebe-se claramente que se a diferença entre o termo e um conceito não for clara, as declarações apresentadas levarão a inferências errôneas na interpretação algorítmica e semântica dos dados. Nessa perspectiva, a necessidade de uma norma como a ISO 25964 se torna evidente no século XXI (Clarke; Zeng, 2012).

A norma ISO 25964-1:2011 *Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval* fornece recomendações para o desenvolvimento e manutenção de tesouros destinados a sistemas de recuperação de informações. É aplicável a vocabulários usados para recuperar recursos informacionais de todos os tipos, independentemente da mídia usada (texto, som, imagem estática ou em movimento, objeto físico ou multimídia), incluindo bases e portais de conhecimento, bancos de dados bibliográficos, museu ou coleções multimídia e os itens nelas contidos (ASTM COMPAS, 2022; International Organization for Standardization, 2011).

A ISO 25964-1:2011 também fornece um modelo de dados e um formato recomendado para a importação e exportação de dados de tesouros; é aplicável a tesouros monolíngues e multilíngues; não é aplicável à preparação de índices finais, embora muitas de suas recomendações possam ser úteis para esse fim; não é aplicável aos bancos de dados ou *software* usados diretamente em aplicativos de pesquisa ou indexação, mas prevê as necessidades de tais aplicativos entre suas recomendações para o gerenciamento de tesouros (ASTM COMPASS, 2022; International Organization for Standardization, 2011).

A norma ISO 25964-2:2013. *Information and documentation — Thesauri and interoperability with other vocabularies — Part 2: Interoperability with other vocabularies* é aplicável a tesouros e outros tipos de vocabulário comumente usados para recuperação de informações. Essa norma descreve, compara e contrasta os elementos e características desses vocabulários que estão envolvidos quando a interoperabilidade é necessária. Oferece recomendações para o estabelecimento e manutenção de mapeamentos entre vários tesouros, ou entre tesouros e outros tipos de vocabulários (ASTM COMPASS, 2022; International Organization for Standardization, 2011).

A publicação da norma ISO 25964 (2011; 2013) atualiza e substitui todas as normas referentes à construção de tesouros e outros vocabulários controlados que vieram antes dela (RAMALHO, 2015), no caso, as 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*. No entanto, as duas partes da norma

apresentada foram publicadas após a divulgação das recomendações SKOS pelo W3C, fazendo com que algumas divergências fossem ocasionadas entre a ISO 25964 e o modelo SKOS (Alvite-Díez; Martínez-González, 2021), pois a versão inicial do modelo de dados SKOS considerou as normas antigas (Barbosa; Viera, 2022; Pastor-Sánchez, 2016).

O *Guia sobre a construção de tesouros* da Agência Nacional de Vigilância Sanitária (ANVISA) (Shintaku, 2021) e o *Manual do planejamento, construção e manutenção do tesouro Unesp para bibliotecas: do conceitual ao praxis* (Fujita; Moreira, 2021) consideram a norma ISO 25964 como uma das principais referências para a apresentação de instruções e conceitos. No entanto, apenas o manual da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP) considera as duas partes da norma, enquanto o guia da ANVISA considera somente a parte 1:2011 (tesouro para recuperação da informação), não abordando aspectos da interoperabilidade entre SOC.

Em relação às alterações que a norma ISO 25964 apresenta em relação às normas anteriores, o manual da UNESP, aponta que na parte 1, as principais alterações foram:

- a) ênfase na distinção entre termo e conceito e a prioridade ao conceito;
- b) fusão na mesma norma do tratamento de tesouros monolíngues e multilíngues, com privilégio aos primeiros;
- c) abordagem prioritária do tesouro em relação a outros tipos de SOC;
- d) ampliação do conjunto de definições dos termos utilizados na norma (glossário);
- e) respeito às diferenças culturais no tratamento das relações associativas, principalmente nos tesouros multilíngues;
- f) tratamento das relações de equivalência com observância das variações linguísticas em diferentes idiomas;
- g) inclusão de seção que trata da análise facetada e de sua importância para a construção de tesouros;
- h) possibilidade de especificação no uso das tags para indicação das relações hierárquicas, com possibilidade de distinção entre relações hierárquicas genéricas, partitivas ou de instanciação;
- i) indicações de critérios para seleção de softwares para gestão de tesouros (Fujita; Moreira, 2021, p. 61).

No que tange à parte 2:2013 (interoperabilidade com outros vocabulários), as principais alterações foram:

- a) inclusão de modelos estruturais para mapeamento entre diferentes vocabulários;
- b) discussão sobre tipos de mapeamento: equivalência, hierárquicos, associativos;
- c) uso de mapeamentos para recuperação da informação;
- d) tratamento da pré-coordenação, gestão e exibição dos mapeamentos;
- e) descrição geral das características de “outros

vocabulários” (expressão presente no título da norma), incluindo para cada um: componentes semânticos e relações conceituais e relações de interoperabilidade e mapeamento com o tesouro. (Fujita; Moreira, 2021, p. 61-62).

Na referida norma, as tags (notações) tradicionais e as novas para estruturação de tesouros não indicam as relações entre termos, como nas normas anteriores, mas indicam as relações entre conceitos. Diante disso, a norma prevê o desenvolvimento de tesouros conceituais.

Ao examinar a primeira parte da norma ISO 25964 e sem esgotar todos os aspectos e modificações, foram identificados vários pontos relevantes. Isso inclui tesouros mencionando outros tipos de vocabulários, a aplicabilidade a recursos além do texto, como som, imagem em movimento e objetos multimídia. Também houve a expansão do número de conceitos abordados no padrão, considerando diferentes relações associativas, mesmo quando um tesouro é utilizado por falantes de diferentes idiomas. O tratamento de relações de equivalência dentro do mesmo idioma e entre diferentes idiomas também foi abordado. O capítulo sobre termos compostos foi alterado para considerar termos complexos, e foram feitas melhorias na simbologia utilizada para indicar relações, permitindo distinguir entre diferentes tipos e aspectos adicionais, como definições e notas históricas sobre os termos, bem como categorias. A norma também oferece recomendações para *software* de gerenciamento de tesouros (Lara, 2013).

Na segunda parte da norma ISO 25964, dedicada ao mapeamento dos tesouros, incluindo outros vocabulários, paralelamente à representação de equivalência dentro do mesmo vocabulário, é recomendado buscar a equivalência entre idiomas, por meio da compreensão de modelos estruturais para o mapeamento entre vocabulários. Esses modelos incluem diferentes tipos de mapeamento, como equivalente, hierárquico e associativo, e abrangem o uso desses mapeamentos para recuperação de informações, o tratamento da pré-coordenação nos mapeamentos, gerenciamento e exibição dos mapeamentos, entre outros aspectos. Além disso, a norma também aborda o mapeamento de outros tipos de vocabulários, além dos tesouros. Para cada tipo de vocabulário, a norma descreve suas características, escopo, papel na recuperação de informações, relacionamentos e componentes semânticos, em comparação com os tesouros. São fornecidas recomendações específicas para o mapeamento entre esses

vocabulários, juntamente com exemplos que não são baseados em tesouros. A norma também destaca os problemas e limitações dos formatos de intercâmbio de dados terminológicos, como *Machine Readable Cataloging* (MARC), Zthes, DD8723-5 e SKOS, indicando a necessidade de ajustes nesses formatos. Em relação ao mapeamento, o padrão pressupõe que os vocabulários devem ser mantidos como entidades separadas, conectadas por meio de seus conceitos correspondentes. No entanto, o anexo do padrão sugere a alternativa de construir um repositório de dados terminológicos a partir do qual diferentes conceitos e vocabulários possam ser extraídos (Lara, 2013).

Bandyopadhyay e Mukhopadhyay (2015), Martínez-González e Alvite-Díez (2019a,c) apontam que a adoção do modelo SKOS pela maior parte da comunidade de tesouros – editores de tesouros e desenvolvedores de *softwares* para construção, manutenção e gerenciamento de tesouros – ocorria de forma abrangente antes da formulação da referida norma, estimulando a publicação da norma ISO 25964. É neste sentido que a contextualização teórica da norma se encontra nesta subseção do presente trabalho, pois a norma se insere num cenário em que o modelo de dados SKOS já estava sendo amplamente utilizado.

Além disso, o modelo de dados SKOS é voltado para a representação de vários tipos de SOC (Zeng; Mayr, 2018), enquanto a norma ISO 25964 é específica para a representação de SOC do tipo tesouro. Embora a parte 2 da referida norma apresente outros tipos de SOC, a intenção é orientar o mapeamento e interoperabilidade dos tesouros com outros tipos de esquemas conceituais. Assim, “desde a publicação da ISO-25964 foi identificada a necessidade de adequar o SKOS à nova norma” (Pastor-Sánchez, 2016, p. [1], tradução livre).

Em suma, os dois fatores supracitados – 1) A norma ISO 25964 surge quando o modelo de dados SKOS já estava sendo amplamente utilizado para construção de tesouros interoperáveis; e 2) As recomendações do modelo de dados SKOS contemplam todos os tipos de SOC, enquanto as recomendações ISO 25964 visam essencialmente os tesouros – indiciam possíveis similaridades entre as recomendações SKOS e as recomendações da norma ISO 25964.

Nesse contexto, mapear, analisar e comparar elementos que compõem o modelo SKOS e a norma ISO 25964 contribui no desenvolvimento de instrumentos

de representação temática como os tesouros conceituais. Ademais, os padrões e formatos de interoperabilidade são essenciais para o aperfeiçoamento de ambientes semânticos na Web, assim como para a sobrevivência de tesouros já construídos, garantindo a disseminação, compartilhamento e expansão destes e outros tipos de SOC no ambiente Web.

## 5 ANÁLISE COMPARATIVA ENTRE A NORMA ISO 25964 E O MODELO SKOS

A partir da revisão da literatura recuperada nas bases de dados citadas na subseção 2.2, estudo da norma ISO 25964 e recomendações SKOS, discute-se nesta seção os resultados da pesquisa acerca das similaridades e equivalências entre o modelo de dados SKOS e a norma ISO 25964, parte 1:2011 e 2:2013, para representação de tesouros. Na subseção 5.1 encontra-se a descrição do *corpus* analisado. Assim, diante das construções que um tesouro pode ter, os resultados são discutidos de modo a perseguir elementos que caracterizam as seguintes categorias de análise de conteúdo delineadas durante esta pesquisa: relação dos tesouros com outros SOC (subseção 5.2); conceitos e termos (subseção 5.3); domínios, microtesouro e grupos de conceitos (subseção 5.4); relações de equivalência (subseção 5.5); equivalência composta (5.6); relações hierárquicas (subseção 5.7); Principais termos (*topTerms*) e principais conceitos (*topConcepts*) (subseção 5.8); relações associativas (subseção 5.9); notas (subseção 5.10); arranjos (subseção 5.11); e abordagens de aprimoramento, extensões SKOS e usos (subseção 5.12<sup>17</sup>) do modelo de dados SKOS e da norma ISO 25964.

### 5.1 Apresentação do *corpus* analisado

Conforme apontado na subseção 2.2, foram estudadas 21 publicações levantadas a partir da busca em bases de dados nacionais e internacionais, a norma ISO 25964<sup>18</sup>, parte 1:2011 (tesouros para recuperação da informação) e parte 2:2013 (mapeamento com outros vocabulários), e as recomendações para o modelo SKOS<sup>19</sup>, totalizando um *corpus* de 25 documentos. Contudo, torna-se válido

---

<sup>17</sup> Subseção estabelecida em decorrência de o modelo de dados SKOS não contemplar todas as recomendações da norma ISO 25964.

<sup>18</sup> INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 25964: **Information and documentation**: thesauri and interoperability with other vocabularies - part 2: interoperability with other vocabularies. Genebra, 2013.

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

<sup>19</sup> ISAAC, Antoine; SUMMERS, E. (ed.). **SKOS Simple Knowledge Organization System Primer**. W3C Working Group Note, 18 August 2009. Disponível em: <https://www.w3.org/TR/skos-primer/>. Acesso em: 19 set 2019.

MILES, Alistair; BECHHOFER, Sean. **SKOS Simple Knowledge Organization System Reference**. W3C Recommendation, 2009. Disponível em: <http://www.w3.org/TR/skos-reference/>. Acesso em: 20 jun 2022.

ressaltar que, devido à abordagem metodológica quadripolar, alguns autores citados nas Seções 1, 2, 3 e 4 poderão ser mencionados na Seção 5.

O Quadro 8 apresenta a bibliografia analisada de acordo com as respectivas propostas e grupo de abordagens e aplicações que envolvem a relação entre a norma ISO 25964 e o modelo de dados SKOS, em ordem cronológica anual crescente.

**Quadro 8 - Bibliografia analisada e respectivas propostas e grupos**

Autoria	Título	Ano	Propostas das publicações	Grupo de abordagens e aplicações
ISAAC, Antoine; Ed Summers,	SKOS Simple Knowledge Organization System Primer	2009	Guiar e informar sobre o uso do SKOS; ser um complemento ao SKOS Reference, que fornece a referência normativa sobre o SKOS	c) Recomendações, normalização, extensões e usos
MILES, Alistair; BECHHOFFER; Sean	SKOS Simple Knowledge Organization System Reference	2009	Documentar a especificação normativa do SKOS. Destina-se a leitores envolvidos no projeto e implementação de sistemas de informação e que já possuem um bom entendimento da tecnologia da Web Semântica, especialmente RDF e OWL	c) Recomendações, normalização, extensões e usos
INTERNATIONAL ORGANIZATION FOR STANDARDIZATION	ISO 25964: Information and documentation: thesauri and interoperability with other vocabularies - part 1: thesauri for information retrieval	2011	Incluir recomendações para formatos e protocolos de intercâmbio; abordar o desenvolvimento e a manutenção de tesouros, em vez de como utilizá-los na indexação.	c) Recomendações, normalização, extensões e usos
INTERNATIONAL ORGANIZATION FOR STANDARDIZATION	ISO 25964: Information and documentation: thesauri and interoperability with other vocabularies - part 2: interoperability with other vocabularies	2013	Fornecer uma breve descrição informativa das características-chave de outros vocabulários, contrastando seus componentes semânticos com os de um tesouro, para fornecer contexto para os requisitos e recomendações de interoperabilidade, não fornecendo orientações sobre a construção de vocabulários diferentes dos tesouros.  Lidar com os princípios e as práticas da interoperabilidade, especialmente o mapeamento, que se aplicam à maioria dos vocabulários, especialmente os tesouros; fornecer orientações adicionais para lidar com as classes pré-coordenadas que ocorrem em esquemas de classificação e outros vocabulários que utilizam a abordagem de classificação.	c) Recomendações, normalização, extensões e usos
CLARKE, Stella G. Dextre; ZENG, Marcia Lei	From ISO 2788 to ISO 25964: The Evolution of Thesaurus Standards towards	2012	Apresentar a evolução normativa para representação de tesouros.	a) Fundamentação histórico-conceitual e c) Recomendações,

	Interoperability and Data Modeling			normalização, extensões e usos
ISO TC46/SC9/WG8; ISAAC, Antoine	Correspondence between ISO 25964 and SKOS/SKOS-XL Models Developed	2013	Corrigir e/ou atualizar o Apêndice "Correspondências entre as construções ISO-2788/5964 e SKOS" do SKOS Simple Knowledge Organization System Primer. Esta atualização foi necessária porque a norma ISO 25964-1:2011 foi publicada, substituindo as normas anteriores ISO 2788:1986 e ISO 5964:1985.	c) Recomendações, normalização, extensões e usos
WILL, Leonard	The ISO 25964 Data Model for the Structure of an Information Retrieval Thesaurus	2012	Analisar a norma ISO 25964 (parte 1) frente ao modelo SKOS.	c) Recomendações, normalização, extensões e usos
CARACCIOLO, Caterina <i>et al.</i>	The AGROVOC Linked Dataset	2012	Fornecer um breve resumo histórico do tesauro AGROVOC e detalhar sua especificação como um Conjunto de Dados Vinculados na nuvem no contexto do movimento LOD.	a) Fundamentação histórico-conceitual e c) Recomendações, normalização, extensões e usos
PASTOR-SÁNCHEZ, Juan-Antonio	ISO-THES: Ampliando Skos a partir de la norma de tesauros ISO 25964	2012	Analisar a extensão ISO-THES frente à norma ISO 25964	c) Recomendações, normalização, extensões e usos
LARA, Marilda Lopes Ginez de	Documentary languages and knowledge organization systems in the context of the semantic web	2013	Discutir a necessidade de linguagens documentárias formais como condição para seu funcionamento na Web Semântica	a) Fundamentação histórico-conceitual;
BALKAN, Lorna; BELL, Lucy	Linking Thesauri - ELSST as a Hub for Social Science Data Terms	2014	Apresentar SOC do tipo tesauro de arquivo baseado em SKOS e na norma ISO 25964.	b) Técnicas de implementação, manutenção e gerenciamento de tesauros; e c) Recomendações, normalização, extensões e usos
BANDYOPADHYAY, Subarna; MUKHOPADHYAY, Parthasarathi	Developing SKOS- Compliant Multilingual Thesaurus: An ISO 25964 Based Approach	2015	Discutir o processo de projetar um tesauro multilíngue usando Tematres com base no padrão ISO 25964 (I e II) e SKOS; investigar as características das normas ISO 25964 como uma ferramenta de construção de tesauros; construir um tesauro	a) Fundamentação histórico-conceitual; b) Técnicas de implementação, manutenção e gerenciamento de tesauros; e c) Recomendações,

			Multilíngue por Tematres de acordo com as diretrizes da ISO 25964; explorar a eficácia da plataforma Tematres na gestão de dados de tesouros.	normalização, extensões e usos
STELLATO, Armando <i>et al.</i>	Collaborative Development of Multilingual Thesauri with VocBench (System Description and Demonstrator)	2015	Descrever as principais características do VocBench, um aplicativo web de código aberto para edição de tesouros SKOS e SKOS-XL, com um forte foco na colaboração, apoiado pela gestão de fluxo de trabalho para validação e publicação de conteúdo.	b) Técnicas de implementação, manutenção e gerenciamento de tesouros; e c) Recomendações, normalização, extensões e usos
STELLATO, Armando <i>et al.</i>	VocBench: a Web Application for Collaborative Development of Multilingual Thesauri	2015	Apresentar o VocBench, um aplicativo web de código aberto para edição de tesouros que estão em conformidade com os padrões SKOS e SKOS-XL; discutir minuciosamente as principais características do VocBench, detalhando sua arquitetura e o avaliando tanto sob o aspecto funcional quanto de apreciação do usuário; fornecer insights sobre desenvolvimentos futuros.	b) Técnicas de implementação, manutenção e gerenciamento de tesouros; e c) Recomendações, normalização, extensões e usos
ALEXIEV, Vladimir; ISAAC, Antoine; LINDENTHAL, Jutta	On the composition of ISO 25964 hierarchical relations (BTG, BTP, BTI)	2016	Definir propriedades estendidas (BTGE, BTPE, BTIE) e analisar quais composições das propriedades originais de "um passo" e as propriedades estendidas são apropriadas.	b) Técnicas de implementação, manutenção e gerenciamento de tesouros; e c) Recomendações, normalização, extensões e usos
PASTOR-SÁNCHEZ, Juan-Antonio	Proposal to represent the UNESCO Thesaurus for the semantic web applying ISO-25964	2016	Mostrar como foi aplicado o padrão ISO-25964 para representar o tesouro da UNESCO por meio de tecnologias da web semântica; analisar o uso conjunto das ontologias SKOS e ISO-THES para representar tesouros de acordo com o modelo de dados da norma ISO-25964.	b) Técnicas de implementação, manutenção e gerenciamento de tesouros; e c) Recomendações, normalização, extensões e usos
BERTI JUNIOR, Decio Wey <i>et al.</i>	Semiautomatização de relações em tesouros: uma proposta para o refinamento de relacionamentos semânticos a partir do tesouro agrovoc	2017	Apresentar uma análise quantitativa do refinamento da estrutura semântica do Tesouro AGROVOC, visando um modelo semiautomatizado para refinamento de relacionamentos semânticos em tesouros.	b) Técnicas de implementação, manutenção e gerenciamento de tesouros; e c) Recomendações, normalização, extensões e usos
ZENG, Marcia Lei; MAYR, Philipp	Knowledge Organization Systems (KOS) in the Semantic	2018	Fornecer uma visão geral do que o movimento LOD KOS trouxe para várias comunidades e usuários; examinar um	a) Fundamentação histórico-conceitual; b)

	Web: a multi-dimensional review		conjunto de casos coletados (experimentais ou em aplicações reais); encontrar os usos do LOD KOS para compartilhar práticas e ideias entre comunidades e usuários; examinar as funções do LOD KOS em várias dimensões.	Técnicas de implementação, manutenção e gerenciamento de tesouros; e c) Recomendações, normalização, extensões e usos
MARTÍNEZ-GONZÁLEZ, M Mercedes; ALVITE-DÍEZ, María-Luisa;	A semantic web methodological framework to evaluate the support of integrity in thesaurus tools	2019	Analisar a integridade dos tesouros e a evolução da ISO 2788 para a ISO 25964; apresentar uma proposta metodológica para avaliação do suporte à integridade em ferramentas tesouros, decorrente dos resultados obtidos.	a) Fundamentação histórico-conceitual; b) Técnicas de implementação, manutenção e gerenciamento de tesouros; e c) Recomendações, normalização, extensões e usos
MARTÍNEZ-GONZÁLEZ, M Mercedes; ALVITE-DÍEZ, María-Luisa	Thesauri and Semantic Web: Discussion of the Evolution of Thesauri Toward Their Integration With the Semantic Web	2019	Comparar e analisar elementos e estruturas presentes no padrão de tesouros ISO 25964 e no SKOS, o padrão da Web Semântica para representação de SOC. Além disso, contrastar as regras de integridade dos tesouros com os axiomas do SKOS.	c) Recomendações, normalização, extensões e usos
MARTÍNEZ-GONZÁLEZ, M Mercedes; ALVITE-DÍEZ, María-Luisa	The support of constructs in thesaurus tools from a Semantic Web perspective: Framework to assess standard conformance	2019	Estudar a evolução da ISO 2788 para a ISO 25964 e as relações entre os construtos nas normas ISO 2788/ISO 25964 e o SKOS; propor um <i>framework</i> metodológico, que se concentra no suporte aos construtos, para avaliar a qualidade de conformidade das ferramentas de gerenciamento de tesouros.	a) Fundamentação histórico-conceitual; b) Técnicas de implementação, manutenção e gerenciamento de tesouros; e c) Recomendações, normalização, extensões e usos
SALIBA, Elie Mario <i>et al.</i>	Zoonom: Gathering the concepts of zoological nomenclature in an electronic thesaurus	2021	Apresentar o tesouro Zoonom, baseado em SKOS e ISO 25964.	b) Técnicas de implementação, manutenção e gerenciamento de tesouros; e c) Recomendações, normalização, extensões e usos
ALVITE-DÍEZ, María-Luisa; MARTÍNEZ-GONZÁLEZ, Mercedes	Representación de tesouros en SKOS: tendencias y desafíos para su integración en la Web Semántica	2021	Analisar e contrastar os construtos e regras de integridade da ISO 25964 e do SKOS; examinar as representações em SKOS dos tesouros AGROVOC, EuroVoc e Unesco, a fim de estudar as soluções adotadas.	b) Técnicas de implementação, manutenção e gerenciamento de tesouros; e c) Recomendações,

				normalização, extensões e usos
RIO-BRANCO, Luciana Beatriz Piovezan; MOREIRA, Walter; FUJITA, Mariângela Spotti Lopes	Mapeamento entre sistemas de organização do conhecimento: revisão sistemática de literatura	2021	Compreender as funções dos mapeamentos entre SOC desenvolvidos na norma ISO 25964-2 para o estabelecimento da interoperabilidade semântica nesses sistemas.	a) Fundamentação histórico-conceitual; b) Técnicas de implementação, manutenção e gerenciamento de tesouros; e c) Recomendações, normalização, extensões e usos
BARBOSA, Everton Rodrigues; VIERA, Angel Freddy Godoy.	Relações semânticas e interoperabilidade em tesouros representados em skos.	2022	Compreender como o modelo de dados Simple Knowledge Organization System e seus modelos de extensão têm sido utilizados para promover a interoperabilidade com outros vocabulários e refinar as relações semânticas em tesouros na web.	a) Fundamentação histórico-conceitual; b) Técnicas de implementação, manutenção e gerenciamento de tesouros; e c) Recomendações, normalização, extensões e usos

Fonte: Elaborado pela autora.

Conforme mencionado na subseção 2.2, a expressão “abordagens” se refere aos aspectos de assunto com os quais o conteúdo da publicação mais se aproxima, enquanto a expressão “aplicações” remete a possibilidades de utilização prática do conteúdo da publicação. Neste sentido, os grupos de abordagens e aplicações são compostos de acordo com os elementos que mais se destacaram e, portanto, permitiram contextualizar as publicações em prol da facilitação da análise de conteúdo categorial, pois, de um modo geral, as publicações poderiam se encaixar nos três grupos apresentados.

Desta forma, as propostas das publicações identificadas se constituem num sentido mais amplo que os objetivos, pois contextualizaram a posterior leitura integral do documento recuperado. Assim, de acordo com o Quadro 8, uma publicação pode pertencer a mais de um grupo de abordagens e aplicações que envolvem a relação entre a norma ISO 25964 e o modelo de dados SKOS.

Ademais, para compreender o cenário comparativo entre o modelo de dados SKOS e a norma ISO 25964 é preciso considerar que

Para garantir uma boa correspondência entre as recomendações desses padrões complementares, as equipes responsáveis por eles mantiveram contato constante. Os modelos de dados respectivos não são idênticos, porque a ISO 25964 deve atender às necessidades de todos os tipos de tesouros (seja para uso na Web ou para outras aplicações), enquanto o SKOS deve abranger todos os tipos de SOC (incluindo esquemas de classificação e muitos outros que não estão em conformidade com a ISO 25964). Apesar das diferenças, no entanto, há uma boa harmonização, o que possibilita o desenvolvimento de correspondências entre os componentes dos modelos de dados. (National Information Standards Organization, 2023, p. [1], tradução livre)

Desta forma, percebe-se que tanto a norma ISO 25964 quanto o modelo SKOS são considerados padrões pelas equipes responsáveis.

## **5.2 Tesouro em relação a outros SOC**

Nesta subseção, o termo tesouro será identificado em diferenciação a outros SOC de acordo com as comparações que mais se destacaram no *corpus*: comparação entre tesouros e ontologias e entre tesouros e esquemas de classificação. Visto que “uma vez que a mera codificação do SOC em SKOS não resolve os problemas de interoperabilidade” (Lara, 2013, p. 148, tradução livre), a

perspectiva conceitual acerca dos tesouros apresentada a partir da bibliografia que relaciona a norma e as recomendações é fundamental para a detecção de similaridades.

Se tanto o indexador quanto o pesquisador selecionarem o mesmo termo para representar um conceito, documentos relevantes serão encontrados. Esse princípio fundamental é a base do *design* de um tesouro, embora também possa ser aplicado em situações em que computadores realizem as escolhas (International Organization for Standardization, 2011).

Os tesouros, enquanto linguagens documentárias pós-coordenadas estruturadas como vocabulários controlados para controle terminológico, cumprem duas importantes funções no contexto de recuperação da informação: apoio à indexação (atividade interna de SRI) e orientação ao usuário informacional sobre os pontos de acesso assunto que podem ser empregados para pesquisa no SRI (atividade externa de SRI). O ponto máximo de sucesso de um tesouro ocorre quando, tanto o indexador quanto o pesquisador, são orientados a expressar o mesmo conceito usando o mesmo termo – orientação que decorre do controle de vocabulário do domínio de conhecimento representado pelo tesouro (International Organization For Standardization, 2011). Além disso, os tesouros devem possuir três características essenciais: a) auxiliar a compreensão intuitiva dos usuários comuns; b) possibilitar inferências por parte das máquinas; e c) permitir a eficiente troca de dados (interoperabilidade) (Bandyopadhyay; Mukhopadhyay, 2015). Neste sentido, a estrutura de um tesouro estruturado de acordo com um modelo de dados também permitirá a expansão das relações semânticas entre os termos, possibilitando a coordenação de termos durante a formulação das estratégias de busca (Alexiev; Isaac; Lindenthal, 2016).

A recomendação *SKOS Primer* (Isaac; Summers, 2009) não apresenta definição específica para tesouro, embora cite este SOC para formular diversos exemplos para entendimento de uso do modelo. Por outro lado, a recomendação *SKOS Reference* (Miles; Bechhofer, 2009) define a ideia de tesouro a partir da explicação da diferença entre um tesouro e um modelo de dados, mencionando os SOC ontologia e sistema de classificação. A definição apresentada considera que o "conhecimento" expresso de forma explícita em uma ontologia formal consiste em

conjuntos de axiomas e fatos. Por outro lado, um tesauro ou esquema de classificação possui uma natureza completamente diferente, não apresentando axiomas ou fatos declarados. Em vez disso, um tesauro ou esquema de classificação é uma ferramenta que identifica e descreve ideias distintas, chamadas "conceitos", usando linguagem natural e meios informais. Esses conceitos podem ser organizados em hierarquias e redes de associação para mapear um domínio específico. No entanto, é importante destacar que essas estruturas não têm uma semântica formal e não devem ser consideradas como axiomas formais ou fatos sobre o mundo. Sua principal finalidade é oferecer um mapa prático e intuitivo do domínio, auxiliando na organização e localização de objetos relevantes, como documentos (Miles; Bechhofer, 2009). Neste sentido, é possível perceber que a recomendação *SKOS Reference* aproxima as estruturas de tesouros e esquemas de classificação. Essa aproximação ocorre em vários pontos da referida recomendação. Miles e Bechhofer (2009) pontuam que tesouros e esquemas de classificação são aplicados em diversas situações, considerando tesouros e esquemas de classificação como SOC semiformais e individuais, pois, antes do surgimento de tecnologias de interoperabilidade, a maioria deles era tratado como sistemas fechados por terem unidades conceituais definidas dentro de um sistema, não podendo participar de outros sistemas, ainda que mapeados para inclusão de unidades conceituais em outros sistemas.

A norma ISO 25964-1 (2011) esclarece que quando os tesouros são implementados em um sistema de busca ou sistema de indexação, os termos e relacionamentos contidos no tesauro podem ser utilizados de maneiras diversas: abrangendo a expansão da busca; apresentando sugestão de termos de busca alternativos; como suporte para aperfeiçoamento da busca por meio de agrupamento ou outros métodos; para detecção de erros ortográficos comuns; e o apoio à indexação automática. Assim, a norma esclarece que “o objetivo de manipular termos é apoiar a recuperação de conceitos” (International Organization for Standardization, 2011, p. 16, tradução livre).

A norma ISO 25964-2 (2013) apresenta as características-chave de oito tipos de SOC: esquemas de classificação, esquemas de classificação usados para gestão de registros (arquivos), taxonomias, esquemas de cabeçalho de assunto, ontologias, terminologias, listas de autoridade de nomes e anéis de sinônimos. Além de abordar

o mapeamento de cada um desses SOC entre tesouros, a norma também fornece esclarecimentos sobre os seus componentes semânticos e relacionamentos comparados aos de um tesouro.

### **5.2.1 Esquemas de classificação**

Um *esquema de classificação* (ou sistema de classificação) é um sistema de classes ou categorias utilizado para organizar recursos de informação. Ele pode ser aplicado tanto na organização física de itens em bibliotecas quanto na organização de registros em catálogos. A abordagem básica da classificação envolve a organização de conceitos em classes, que são subdivididas em classes mais específicas, conforme necessário. Cada classe possui um título e, possivelmente, uma descrição mais detalhada que delimita seu escopo. Esse tipo de descrição funciona como nota de escopo em um tesouro, listando os tópicos incluídos e fornecendo referências cruzadas para outras classes relacionadas.

Em um esquema de classificação de assuntos, cada classe é identificada por uma notação, geralmente uma sequência curta de caracteres alfabéticos, numéricos ou símbolos, que representam a classe. A abordagem de classificação geralmente permite a pré-coordenação de conceitos e é adotada em diferentes tipos de vocabulários, incluindo taxonomias e esquemas usados em sistemas de gestão de registros (International Organization for Standardization, 2013).

Os esquemas de classificação enumerativos apresentam todas as possíveis sequências de notação disponíveis para os usuários, enquanto os esquemas sintéticos e facetados fornecem regras que permitem aos usuários criar sequências de notação para combinações de conceitos não previamente enumeradas. Esses esquemas desempenham um papel importante na organização e recuperação de informações em diversos contextos (International Organization for Standardization, 2013), 2. Exemplos de notação são apresentados na norma, sendo um deles pertencente à Classificação Decimal Universal (CDU), idealizada por Paul Otlet a partir da Classificação Decimal de Dewey (CDD).

Ao tratarem da “razão de ser do tesouro”, Clarke e Zeng (2012) apontam que o tesouro é uma ferramenta intermediária que apoia o acesso à informação, encontrando-se entre os esquemas de classificação e as formas de pesquisa em

texto completo. Diferentes ferramentas e abordagens foram utilizadas nesse contexto, mas o tesauro se posiciona como uma opção que equilibra essas duas extremidades do espectro. Isto porque a abordagem de esquemas de classificação depende de um esquema de conhecimento prévio em um domínio específico, onde cada assunto recebe um código único. Isso facilita a busca por documentos relevantes, mas pode aumentar os custos e não é popular entre os usuários. Por outro lado, a pesquisa de texto completo permite que os usuários procurem por palavras-chave em qualquer lugar dos documentos, mas tem suas limitações uma vez que um assunto pode ser expresso de várias maneiras. Uma alternativa é usar um tesauro para orientar os usuários a usarem termos consistentes para conceitos específicos, permitindo que pesquisem com palavras comuns. O fundamental é fornecer métodos práticos para representar conceitos e suas inter-relações de forma clara e simples (Clarke; Zeng, 2012).

Caracciolo *et al.* (2012) aponta que o Tesauro AGROVOC é publicado como um conjunto de dados LOD e, entre outros vocabulários, está vinculado a uma versão experimental da CDD<sup>20</sup> em formato *Linked Data*.

### **5.2.2 Esquemas de classificação para a gestão de registros**

Muitas organizações utilizam esquemas de classificação personalizados para gerenciar seus registros oficiais organizados em arquivo. Esses esquemas, denominados *esquemas de classificação para a gestão de registros*, diferem dos esquemas usados em bibliotecas, como tesouros e sistemas de classificação, por exemplo, pois estão mais preocupados com o contexto empresarial e as atividades específicas da organização. Um esquema de classificação bem projetado leva em consideração as atividades da organização, os requisitos regulatórios, os fatores de risco e as necessidades de segurança e responsabilidade. Ele é utilizado para associar registros às funções, atividades e transações comerciais correspondentes, mantendo-os organizados em uma sequência cronológica correta (International Organization for Standardization, 2013).

Os esquemas de classificação para gestão de registros são desenvolvidos para atender a requisitos legais relacionados à manutenção de registros precisos e

---

<sup>20</sup> <http://dewey.info/>

completos. A retenção e o descarte adequados são fatores-chave nesse processo, pois é importante que o conteúdo dos arquivos e pastas possa ser retido pelo tempo necessário e descartado conforme um cronograma pré-determinado. Além disso, a necessidade de controlar o acesso aos registros pode influenciar a forma como os arquivos são agrupados e organizados. Um exemplo comum de esquema de classificação funcional usado na gestão de registros é composto por classes de alto nível que são subdivididas em níveis adicionais, conforme necessário. Essa estrutura permite uma organização eficiente dos registros, facilitando o acesso, a retenção e o descarte adequados (International Organization for Standardization, 2013).

Além da norma ISO 25964-2:2013, nenhum outro texto do *corpus* de análise abordou os esquemas de classificação para a gestão de registros. Em atenção aos usos possíveis para este tipo de SOC em arquivos, percebe-se a validade de considerá-lo para interoperabilidade com os tesauros, que também podem compor ambientes organizacionais, de memória institucional e gestão do conhecimento.

Em suma, enquanto os esquemas de classificação são linguagens documentárias pré-coordenadas, os tesauros são linguagens documentárias pós-coordenadas (Caracciolo *et al.*, 2012; Clarke & Zeng, 2012; International Organization for Standardization, 2013).

### 5.2.3 Ontologia

A norma ISO 25964-2:2013 aponta a variedade de significados atribuídos ao termo "*ontologia*", desde modelos semânticos de dados até análises filosóficas dos conceitos em um domínio. No entanto, a norma adota a definição da Ciência da Computação, sendo *ontologia* uma especificação formal e explícita de uma conceituação compartilhada. Na modelagem de dados, a ontologia é interpretada como o uso de uma linguagem formal para representar um domínio de conhecimento, permitindo a verificação da consistência das afirmações de conhecimento e a inferência de novos fatos (International Organization for Standardization, 2013).

Uma ontologia, juntamente com um conjunto de fatos, forma uma base de conhecimento. A ontologia desempenha um papel fundamental no raciocínio,

permitindo a inferência de pertinência de classes para indivíduos, relacionamentos entre classes e propriedades, e verificação da consistência do conhecimento. No campo médico, por exemplo, uma ontologia de doenças pode ser usada para inferir a doença com base nos sintomas do paciente ou identificar a causa comum de sintomas apresentados por diferentes pacientes (International Organization for Standardization, 2013).

A norma também aborda a possibilidade de reengenharia de um tesauro como uma ontologia de acordo com os objetivos para tal, onde “os conceitos e relacionamentos no tesauro devem ser traduzidos para a ontologia, e as características específicas da ontologia devem ser adicionadas conforme necessário” (International Organization for Standardization, 2013, p. 77, tradução livre).

Will (2012) explica que as estruturas relacionais dos tesouros representados em SKOS precisam ser muito bem definidas para possibilitar a compatibilidade e interoperabilidade com ontologias, principalmente com relação às relações hierárquicas e associativas. Por exemplo, ao especificar a natureza da relação, como causa/efeito, processo/produto ou pessoa/disciplina, permitindo o tratamento de “conceitos relacionados” abrangentes (TR) quando necessário, isso permitirá que um tesauro se aproxime da abordagem adotada em ontologias, onde a natureza de todos os relacionamentos é especificada (Will, 2012).

Ao abordar a versão dois do VocBench<sup>21</sup> como *software* para construção e gerenciamento de ontologias e tesouros em SKOS e SKOS-XL (saber mais na subseção 5.12.2), Stellato *et al.* (2015b) aponta que no desenvolvimento tradicional de ontologias, o raciocínio desempenha um papel crucial na validação formal da ontologia, garantindo a sua consistência lógica. No entanto, essa abordagem não se aplica aos tesouros SKOS. Os tesouros SKOS são estruturas de organização de informações que diferem das ontologias tradicionais. Ao contrário das ontologias,

---

<sup>21</sup> VocBench (também abreviado como VB) foi concebido como uma aplicação web acessível por meio de qualquer navegador moderno, liberando assim os usuários finais da instalação e configuração de software. Muitas das limitações do VB1 em relação aos requisitos descritos na versão anterior estavam relacionadas à falta de um backend real RDF. Enquanto o VB1 era baseado na API do Protégé 3 OWL (uma interface OWL não nativa em torno do modelo baseado em frames do Protégé 3), o VB2 foi redesenhado para depender das capacidades do Semantic Turkey [11], uma plataforma de gerenciamento RDF já desenvolvida e atualmente mantida pela Universidade de Tor Vergata. (STELLATO *et al.*, 2015b, p.41, tradução livre)

que são formalmente representadas por axiomas formais *Ontology Web Language* (OWL), a maioria das suposições sobre o uso do SKOS não é codificada explicitamente como axiomas formais OWL. Em outras palavras, o SKOS não segue as mesmas regras de raciocínio formal usadas em ontologias tradicionais (Stellato *et al.*, 2015b).

Zeng e Mayr (2018) esclarecem que “ontologias leves” são aquelas que utilizam classes e propriedades ontológicas para expressar os SOC convencionais. Essa abordagem é popular entre aqueles que publicam SOC modelados em SKOS. Essas “ontologias leves” geralmente não são consideradas “ontologias de referência” que possuem teorias complexas e axiomáticas com o objetivo de esclarecer os significados pretendidos dos termos usados em domínios específicos. Em vez disso, são consideradas “ontologias de aplicação” que fornecem uma estrutura terminológica mínima para atender às necessidades de uma comunidade específica. Todavia, o termo “ontologias” tem sido aplicado a diversos tipos de vocabulários, enquanto abordagens como ontologias superiores e ontologias híbridas amplamente utilizadas na criação de novos SOC (Zeng; Mayr, 2018). Neste cenário, os tesouros são compreendidos como “ontologias leves”.

A substância ontológica dos conceitos em SOC podem ser comparadas com os resultados obtidos (conceitos) ao aplicar técnicas de PLN às representações lexicais (literais RDF) dos conceitos SKOS. Nesse contexto, há comparações entre o *Functional Requirements for Subject Authority Data* (FRSAD) e o *Simple Knowledge Organization System* (SKOS), utilizando técnicas de PLN (Alvite-Díez; Martínez-González, 2021; Martínez-González; Alvite-Díez, 2019b). Essas comparações se relacionam a abordagens centradas nos tesouros modelados em SKOS, identificando suas fortalezas e oportunidades para o avanço em direção a soluções voltadas à interoperabilidade com ontologias formais e outros SOC no contexto de IA.

Ao falar sobre a Web Semântica, ou ambientes semânticos na Web, é importante mencionar as ontologias, que são ferramentas conceituais poderosas para compartilhar conhecimento. Existem ontologias gerais, abrangendo várias áreas, e ontologias específicas para domínios como geoinformação e jurídico. As ontologias permitem descrever e relacionar recursos com base em propriedades

compartilhadas. Os tesouros também são usados para compartilhar conhecimento, e sua comparação com as ontologias tem sido discutida há anos.

O padrão SKOS foi desenvolvido para representar SOC na Web Semântica. Já as ontologias lidam com classes e instâncias (indivíduos), permitindo o raciocínio por herança de propriedades, enquanto os tesouros lidam apenas com conceitos<sup>22</sup>. As diferenças entre ontologias e tesouros justificam a existência do SKOS. Ontologias e tesouros têm sido usados juntos, com tesouros enriquecendo ontologias a partir de informações terminológicas, enquanto o SKOS facilita a integração dos tesouros na Web Semântica. Além disso, tesouros têm sido representados com SKOS, e ontologias específicas foram criadas para acompanhar os tesouros, estendendo o SKOS de acordo com a semântica específica de cada tesouro (Martínez-González; Alvite-Díez, 2019b).

Martínez-González e Alvite-Díez (2019c) afirmam que assim como a norma ISO 25964 e as recomendações SKOS podem ser comparadas entre si, tesouros também podem ser comparados a ontologias. Sobre isso, os autores defendem que tesouros e ontologias são SOC utilizados para organizar o conhecimento de uma comunidade. Apesar dos esforços para obter ontologias a partir de tesouros e outros SOC, existem diferenças relevantes entre eles. Essas diferenças justificam a existência do padrão SKOS, projetado para representar SOC na Web Semântica. Assim, trabalhos específicos com tesouros são interessantes, especialmente quando se trata de tesouros consolidados que dependem do consenso de amplas comunidades (Martínez-González; Alvite-Díez, 2019c).

O uso de ontologias é compatível com a manutenção de tesouros. Por exemplo, o Serviço de Publicações da União Europeia utiliza ontologias jurídicas juntamente com o Tesouro EuroVoc em seus sistemas de informação, facilitando os processos de recuperação da informação (Martínez-González; Alvite-Díez, 2019c).

---

<sup>22</sup> O raciocínio em ontologias permite que indivíduos ou instâncias herdem as propriedades das classes a que pertencem. No entanto, nos tesouros, que tratam de conceitos, esse tipo de raciocínio não se aplica. Por exemplo, em uma ontologia, "Terra" é uma instância da classe "Planeta" e herda as propriedades de um planeta, enquanto em um dicionário de sinônimos, "Terra" e "Planeta" são conceitos. A ISO 25964-2:2013 fornece um exemplo para ilustrar essa diferença. Em um tesouro, "Terra" e "Planeta" podem ser dois conceitos relacionados hierarquicamente. Em uma ontologia sobre mecânica celeste, "Planeta" seria estabelecido como uma classe e "Terra" como um indivíduo único no domínio. Essa distinção permite que o indivíduo herde todas as propriedades das classes a que pertence (Martínez-González; Alvite-Díez, 2019c).

Percebe-se o quanto as ontologias e o tesauros estão relacionados, tanto conceitualmente, quanto funcionalmente no contexto de ambientes semânticos na Web. Portanto, observar e atender às normatizações e recomendações vigentes corrobora com a revitalização da representação temática da informação no contexto Web por meio de tesauros conceituais e interoperáveis inter-tesauros, intra-tesauros, entre sistemas, humanos e outros SOC como as ontologias, que, de acordo com a revisão realizada neste trabalho, se aproxima de tecnologias cada vez mais próximas da realidade de uso em rede da linguagem natural e artificial.

### 5.3 Conceitos e termos

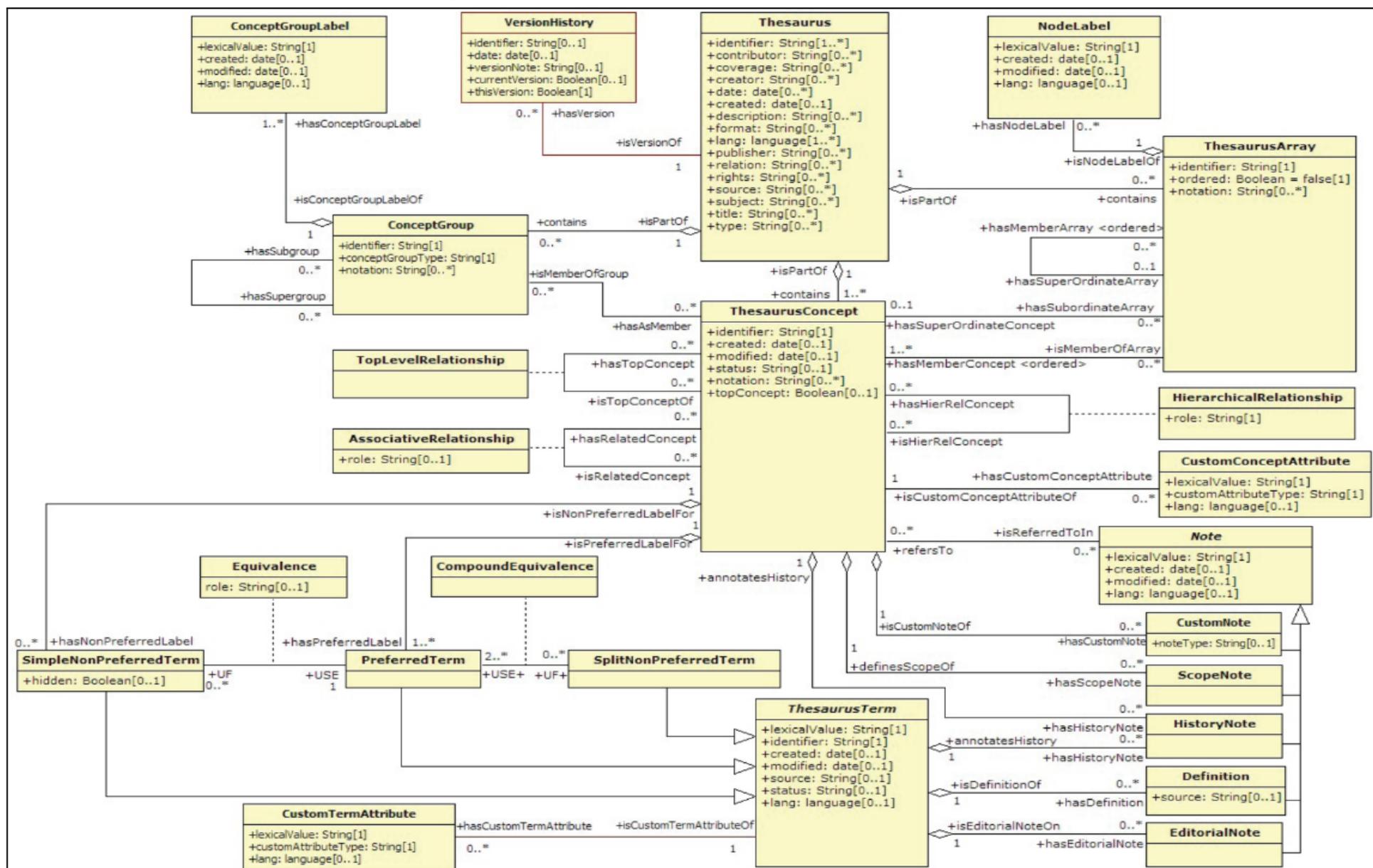
No tesouro, a relação entre termos e conceitos é fundamental. Para Clarke e Zeng (2012, p. 21, tradução livre) “a única maneira de um tesouro listar conceitos em ordem alfabética é representá-los como termos”. Nesse sentido, as autoras indicam que o objetivo claro de organização dos conceitos e suas inter-relações não foram inventadas pela norma ISO 2788:1974 *Documentation - Guidelines for the establishment and development of monolingual thesauri*, nem mesmo pelos padrões que a sucederam, ao passo que as tags (ou notações) BT, NT e RT já tinham sido usadas ao longo da década de 60, especialmente no *Thesaurus of Engineering and Scientific Terms* (TEST).

Clarke e Zeng (2012) colocam a ISO 25964 como um modelo de dados, norma e padrão mais rigorosa ao diferenciar explicitamente termos e conceitos. A norma mantém as tags BT, NT e RT, pois já são utilizadas em milhares de tesauros, porém esclarece que elas devem fazer referência a conceitos e não a termos. Isto reforça que a estrutura dos tesauros interoperáveis semanticamente precisa ser baseada em conceitos, não em termos (WILL, 2012). Assim, o padrão define cinco classes básicas: *Thesaurus*, *ThesaurusArray*, *ThesaurusConcept*, *ThesaurusTerm* e *Note*. Os atributos e associações das classes refletem os recursos do tesouro. O padrão é acompanhado por notas explicativas claras, onde uma delas orienta que a representação de conceitos deve ocorrer por meio de termos preferenciais e não preferenciais, esclarecendo que “cada conceito no tesouro é representado por um termo preferencial por idioma e por qualquer número de termos não preferenciais” (International Organization for Standardization, 2011, p. 105, tradução livre).

A norma ISO 25964-1:2011 descreve os tesauros e suas relações entre conceitos e termos. Os conceitos são unidades de pensamento, enquanto os termos são os rótulos utilizados para identificar esses conceitos (WILL, 2012), assim como nas recomendações SKOS. Cada conceito possui um termo preferido em cada idioma, mas também pode ter termos não preferenciais de forma ilimitada. Essa abordagem permite expressar relações clássicas de equivalência entre termos (USE/ USE FOR) e oferece um atributo de "função" para especificar a natureza do relacionamento, por exemplo, o relacionamento entre um termo preferencial e não preferencial pode ter a "função" de "abreviação/forma completa, formal/informal, obsoleto/atual ou científico/popular. Foi considerado desnecessariamente complicado fornecer tais relações entre um termo não preferido e outro" (Will, 2012, p.48-49, tradução livre).

A nova norma ISO 25964 está muito mais próxima do modelo SKOS do que as normas anteriores por várias razões: ela define um modelo de dados que facilita o desenvolvimento de aplicativos de computador e sua aplicação em processos de recuperação de informação, conforme mostrado na Figura 5; esse modelo estabelece um nível conceitual e um nível terminológico (léxico); além de conceitos e termos, são definidas várias classes de estruturas superiores para representar tesauros ("*Thesaurus*"), grupos de conceitos ("*ConceptGroup*") e listas estruturadas de conceitos ("*ThesaurusArray*"); os elementos de anotação são praticamente idênticos aos do SKOS (ISO TC46/SC9/WG8; Isaac, 2012; Pastor-Sánchez, 2013a).

Figura 5 - Modelo de dados definido pela ISO 25964.



Fonte: Pastor-Sánchez (2013a).

Observando a Figura 5, percebe-se que a norma ISO 25964 apresenta diferentes níveis de modelagem, enquanto o SKOS se constitui principalmente a nível de modelo de dados. Essa mesma figura também expõe uma vertente mais computacional da norma.

Bandyopadhyay e Mukhopadhyay (2015) esclarecem que, dentre as características únicas da ISO 25964 em relação à ISO 2788:1986 – que regeu a construção de tesauros monolíngues por décadas e ainda existentes –, destaca-se a orientação de que a garantia do usuário deve prevalecer sobre a garantia literária, onde o termo selecionado como preferencial deve ser de uso predominante pela comunidade.

As diretrizes da norma vigente para a construção de tesauros orientam a dar preferência ao termo mais popular entre a comunidade-alvo para representar um conceito, de modo a escolher: a grafia mais aceita; dar preferência aos termos emprestados de outros idiomas se estes forem bem estabelecidos na comunidade; utilizar neologismos, gírias e jargões quando considerados úteis pela comunidade (Bandyopadhyay; Mukhopadhyay, 2015).

Esta perspectiva de garantia do usuário vai ao encontro dos preceitos de Wittgenstein apresentados por Gracioso e Saldanha (2011), que pontuam que a linguagem não deveria ser entendida pela lógica, mas sim a partir de seu uso, compreensão das funções práticas e não apenas pelo significado isolado das palavras, conforme apresentado na Seção 3. Assim, os estudos ao entorno da norma, bem como o texto da própria norma e recomendações SKOS, demonstram estar alinhados às treze teorias contemporâneas da Ciência da Informação definidas por Araújo (2017), em que todas se aproximam do paradigma social da informação defendido por Capurro (2003), que por sua vez, também é influenciado pela corrente epistemológica hermenêutica.

A norma ISO 25964 ainda esclarece sobre a seleção de termos preferenciais, a partir das seguintes recomendações: usar o nome comum em vez do nome comercial para evitar problemas legais; preferir nomes populares em tesauros gerais e nomes científicos em tesauros científicos. Utilizar nomes genéricos, nomes geográficos (tanto populares quanto oficiais) e nomes de instituições e pessoas corretos e atualizados (Bandyopadhyay; Mukhopadhyay, 2015).

Bandyopadhyay e Mukhopadhyay (2015) explicam que, a norma ISO 25964 enquanto um modelo de dados abstrato é aplicada para representar e utilizar dados em tesouros multilíngues e monolíngues. Se houver diferenças nas estruturas de dados entre os idiomas, é necessário fazer o mapeamento para garantir a consistência. O mapeamento estabelece relações entre os conceitos de diferentes vocabulários, promovendo a interoperabilidade entre sistemas ao permitir a troca e utilização de informações de forma padronizada, aumentando a eficiência das ferramentas de vocabulário. Neste sentido, o SKOS, além de ser um modelo de dados, se converge em uma ferramenta para aplicação de tal interoperabilidade prevista pela norma, possibilitando que os tesouros sejam mais eficazes para o ambiente Web.

Como nas normas anteriores, a ISO 25964 estabelece condições de integridade para um tesouro, em que “as condições de integridade determinam o que é ou não válido em um tesouro” (Martínez-Gonzalez; Alvite-Díez, 2019a, p. 4, tradução livre). As condições de integridade são afirmações presentes na definição formal de um modelo de dados que permitem verificar se os dados fornecidos são coerentes e estão em conformidade com o modelo estabelecido (Martínez-Gonzalez; Alvite-Díez, 2019a). Assim, regras de integridade também definem o que é ou não válido para um tesouro, tanto representado em SKOS como de acordo com a ISO 25964, ou conforme as similaridades de ambos.

A comparação entre as notações SKOS e ISO 25964 se faz importante para analisar a compatibilidade de ambos os modelos em prol da integridade de tesouros representados de acordo com cada modelo. Martínez-González e Alvite-Díez (2019c), Alvite-Díez e Martínez-González (2021) agrupam conjuntos de notações, classes e propriedades da ISO 25964 para examiná-las detalhadamente de modo a discutir as possíveis relações com as representações SKOS. Os autores constroem tal discussão a partir da análise minuciosa de tesouros representados com SKOS. Nesse contexto, em concomitância com a norma e recomendações SKOS, o Quadro 9 compara classes concernentes a conceitos e termos, conforme segue:

**Quadro 9** - Classes na ISO 25964 e no SKOS para tesouros, conceitos e termos.

Classes ISO 25964	Função ISO	Notação SKOS	Função SKOS
-------------------	------------	--------------	-------------

<i>Thesaurus</i>	designar tesouros	<i>skos:ConceptScheme</i>	designar tesouros e outros SOC
<i>ThesaurusConcept</i>	designar conceito de um tesouro	<i>skos:Concept</i>	designar conceito de tesouros e outros SOC
<i>ThesaurusTerm</i>	designar termo do tesouro	<i>skosxl:Label</i>	designar conceito de tesouros e outros SOC

Fonte: Elaborado pela autora.

Conforme mencionado na Seção 4, as recomendações SKOS e a norma ISO 25964 foram publicadas em momentos diferentes, o que fez com que o SKOS tivesse de ser atualizado por meio da extensão SKOS-XL (mais detalhes na subseção 5.12.2), de modo a atender a norma. Deste modo, a notação *skosxl:Label* surge após essa atualização, possibilitando a representação e relacionamentos a nível lexical (termos) dos SOC do tipo tesouros.

Prosseguindo com a esquematização das similaridades entre as recomendações SKOS e a norma ISO 25964, segue comparação das notações de relacionamentos para tesouros, conceitos e termos.

**Quadro 10** - Relacionamentos para tesouros, conceitos e termos na norma ISO 25964 e no modelo SKOS.

<b>Relacionamento ISO 25964</b>	<b>Função ISO</b>	<b>Propriedade SKOS</b>	<b>Função SKOS</b>
<i>isPreferredLabelFor / hasPreferredLabel</i>	relacionar <i>PreferredTerm</i> / e <i>ThesaurusConcept</i>	<i>skos:prefLabel / skosxl:prefLabel</i>	relacionar rótulos (termos) preferenciais a um conceito / ou a um termo
<i>isNonPreferredLabelFor / hasNonPreferredLabel</i>	relacionar <i>NonPreferredTerm</i> / e <i>ThesaurusConcept</i>	<i>skos:altLabel / skosxl:altLabel</i>	relacionar rótulos (termos) alternativos ou não preferenciais a um conceito / ou a um termo
<i>isPartOf / contains</i>	relacionar <i>ThesaurusConcept</i> / e <i>Thesaurus</i>	<i>skos:inScheme</i>	relacionar <i>skos:Concept</i> a <i>skos:ConceptScheme</i> .

Fonte: Elaborado pela autora.

Segundo o exposto na subseção 4.1.2, as propriedades de etiquetagem do SKOS possibilitam representar relações semânticas a nível lexical, relação entre

conceito e termo. A notação *skosxl:altLabel* advém da extensão SKOS-XL, constituindo-se em propriedade que objetifica uma etiqueta/rótulo SKOS de modo a permitir o relacionamento entre termos e não apenas entre conceitos. Martínez-González e Alvite-Díez (2019c) esclarecem que somente termos e conceitos preferenciais podem participar de relacionamentos hierárquicos e associativos. Esta condição está diretamente relacionada às condições de integridade da ISO 25964.

Outras condições de integridade relacionadas à diferenciação entre termos e conceitos na norma ISO 25964, são: unicidade (não deve haver elementos duplicados no tesouro, onde deve haver apenas um termo preferencial para cada conceito em cada idioma); incompatibilidade de relacionamentos (pois alguns relacionamentos são incompatíveis); proibição de ciclos que envolvam relação hierárquica; existência de relação inversa para NT e BT, por exemplo; e exclusão de relacionamentos referentes a um termo que foi excluído (Martínez-González E Alvite-Díez, 2019c). Essas regras de integridade previstas pela ISO 25964 devem ser consideradas não apenas no âmbito da elaboração de tesouros, mas também nos contextos de *software* de construção, gerenciamento e avaliação de tesouros, de modo a automatizar algumas etapas, agilizando, assim, o trabalho de editores e desenvolvedores de tesouros.

As condições de integridade SKOS para representação de diversos SOC são expressas pela recomendação SKOS *Reference* (Miles; Bechhofer, 2009). Essas condições são expressas por um conjunto de 56 axiomas, para os quais Martínez-González e Alvite-Díez (2019c) apresentam apenas alguns desses axiomas em relação às restrições ISO 25964, verificando que nem todos possuem equivalência com as condições de integridade da norma. Dos axiomas de restrição de integridade SKOS que possuem equivalência com as integridades ISO, cita-se o axioma S14 (um recurso não tem mais de um valor de *skos:prefLabel* por *tag* de idioma) que é equivalente a uma das condições de integridade de unicidade da ISO (apenas um termo preferido para cada conceito é permitido).

As regras de integridade garantem consistência aos tesouros, bem como às funções para as quais são destinados. Martínez-González e Alvite-Díez (2019c) esclarecem que mesmo a extensão SKOS-XL detém limitações para oferecer todo o

potencial da ISO 25964 para a representação de tesouros conceituais interoperáveis.

#### 5.4 Domínios, microtesauro e grupos de conceitos

Os tesouros tendem a ser organizados em campos do conhecimento ou domínios temáticos, também denominados de microtesauros, posto que um tesouro pode ser organizado em vários microtesauros. Porém, a ISO 25964 formula uma nova possibilidade: grupos de conceitos (ou domínios).

Os tesouros podem ter vários níveis de domínio que se incluem um ao outro, caracterizando subdomínios, mas também podem ser divididos em vários microtesauros, que por sua vez não podem se repetir dentro de um mesmo tesouro (International Organization for Standardization, 2011) (Martínez-Gonzalez; Alvite-Díez, 2019a), o que contempla parte das condições de integridade/unicidade previstas pela norma. Nesse sentido, um *ConceptGroup* é “um grupo de conceitos selecionados por algum critério específico, como relevância para uma área temática específica” (International Organization for Standardization, 2011, p. 110, tradução livre). Assim, microtesauros também são considerados *ConceptGroup* pela norma ISO-25964-1:2011.

Muitos tesouros reúnem conceitos em subconjuntos que tendem a ser formados de acordo com as disciplinas, temas e áreas empresariais que representam (International Organization for Standardization, 2011; Will, 2012). De acordo com o orientado pela norma ISO 25964, os conceitos pertencentes a um mesmo grupo não necessariamente necessitam ter relação hierárquica ou associativa entre si e podem ser extraídos de distintas estruturas conceituais ou facetadas do tesouro. Essas facetadas podem se constituir em atividades, pessoas, lugares e coisas. Assim, conceitos aninhados podem ter notações que se diferenciam de outras estruturas, o que possibilita “um arranjo classificado que complementa a hierarquia genérica do próprio tesouro, como em um “Thesaurofacet” ou “Classaurus ” (Will, 2012, p. 51, tradução livre).

Além da classe *ConceptGroup*, a norma ISO 25964, enquanto um modelo de dados, prevê a existência de diferentes tipos de grupos fornecendo as classes *ConceptGroupLabel* (para rotular grupos de conceitos) e *conceptGroupType* (para

rotular tipos específicos de grupos de conceitos) (International Organization for Standardization, 2011). Estas duas últimas notações não possuem equivalentes diretas em SKOS. Ademais, a norma não prevê relação hierárquica entre um *ConceptGroup* e os conceitos que ele contém, onde “cada grupo deve receber um rótulo verbal por idioma” (International Organization for Standardization, 2011, p. 108, tradução livre).

No mesmo sentido, Pastor-Sánchez (2013a, p. 190, tradução livre) afirma que “o SKOS possui duas classes que permitem criar estruturas de agrupamento: esquemas de conceitos e coleções, juntamente com as propriedades correspondentes para relacionar os conceitos a essas estruturas”. Coleções representadas em SKOS (*skos:Collection*) também são grupos de conceitos que não precisam estar relacionadas hierarquicamente. Os conceitos numa *skos:Collection* também podem pertencer a diferentes *skos:ConceptScheme*, o que possibilita o agrupamento de conceitos de diferentes tesouros (Martínez-González; Alvite-Díez, 2019a).

O quadro a seguir explicita a similaridade notacional para designar grupos de conceitos na norma ISO 25964 e modelo SKOS:

**Quadro 11** - Notação para grupos de conceitos na ISO 25964 e SKOS.

Notação ISO 25964	Função ISO	Notação SKOS	Função SKOS
<i>ConceptGroup</i>	indicar um grupo de conceitos, domínio, microtesauro, tema ou subconjunto do tesauro	<i>skos:Collection</i> , <i>skos:ConceptScheme</i>	indicar uma coleção, indicar um tesauro ou outro SOC

Fonte: Elaborado pela autora.

Na sequência, a similaridade notacional e funcional para indicar relacionamentos entre os grupos de conceitos na norma ISO 25964 e modelo SKOS é descrita no quadro 12, a saber:

**Quadro 12** - Notação de relacionamentos para grupos de conceitos na ISO 25964 e no SKOS.

Relacionamento na	Função ISO	Propriedade SKOS	Função SKOS
-------------------	------------	------------------	-------------

ISO 25964			
<i>isConceptGroupLabelOf</i> / <i>hasConceptGroupLabel</i>	relacionar <i>ConceptGroupLabel</i> e <i>ConceptGroup</i>	<i>skos:prefLabel</i> / <i>skosxl:prefLabel</i>	relacionar rótulos (termos) preferenciais a um conceito
<i>isPartOf</i> / <i>contains</i>	relacionar <i>ConceptGroup</i> e <i>Thesaurus</i>	<i>skos:inScheme</i>	relacionar <i>skos:Concept</i> a <i>skos:ConceptScheme</i> .
<i>hasSubGroup</i> / <i>hasSuperGroup</i>	relacionar <i>ConceptGroup</i> e <i>ConceptGroup</i>	<i>skos:member</i>	relacionar duas <i>skos:Collections</i>
<i>isMemberOfGroup</i> / <i>hasAsMember</i>	relacionar <i>ThesaurusConcept</i> e <i>ConceptGroup</i> .	<i>skos:member</i> ; <i>skos:inScheme</i>	relacionar duas <i>skos:Collections</i> ; relacionar <i>skos:Concept</i> a <i>skos:ConceptScheme</i> .

Fonte: Elaborado pela autora.

Uma vantagem de utilizar *skos:Collection* para grupos de conceitos é que o SKOS fornece propriedades que permitem o aninhamento de coleções. A propriedade *skos:member* relaciona duas coleções, enquanto não há uma propriedade no SKOS que relacione dois esquemas de conceitos. Dessa forma, o uso de coleções parece ser mais adequado para o aninhamento de grupos de conceitos em SKOS de diferentes tesouros. No entanto, uma desvantagem é que as coleções não podem ter conceitos principais no SKOS. Enquanto a propriedade *skos:topConceptOf* relaciona um conceito a um esquema de conceitos, ela não se aplica a coleções (Martínez-Gonzalez; Alvite-Díez, 2019a, b). A importância dos conceitos principais será discutida na subseção 5.8.

Percebe-se que as notações, propriedades ou classes do SKOS se repetem na comparação com os elementos similares entre a norma ISO 25964 e o modelo SKOS. Isso ocorre por conta da especificidade da norma para representar tesouros, enquanto o SKOS não necessita de tal especificação, porque abrange diversos tipos de SOC.

A representação de grupos de conceitos com SKOS tem soluções heterogêneas. Esquemas de conceitos e coleções são usados, mas nenhum deles representa exatamente a natureza dos grupos de conceitos em ISO 25964. Por isso, ontologias *ad hoc* que estendem o SKOS propõem novas classes para esses

grupos, bem como para domínios e microtesauros. Essas novas classes preservam sua natureza específica (semântica dos grupos de conceitos em tesauros), que não é encontrada no SKOS (Alvite-Díez; Martínez-González, 2021; Martínez-Gonzalez; Alvite-Díez, 2019b;).

A natureza dos grupos de conceitos definidas pela ISO 25964 não é representada pela propriedade *skos:Collection*, pois a recomendação SKOS *Primer* define coleções de conceitos como uma estrutura parecida com SOC do tipo anéis de sinônimos, conforme exemplo a seguir:

```

leite
  <leite por fonte animal>
    leite de vaca
    leite de cabra
    leite de búfalo (Isaac; Summers, 2009, p. [1], tradução livre)

```

Tal constatação reforça a ideia de que há apenas similaridade entre *skos:Collection* e *ConceptGroup* e não equivalência exata.

## 5.5 Relações de equivalência

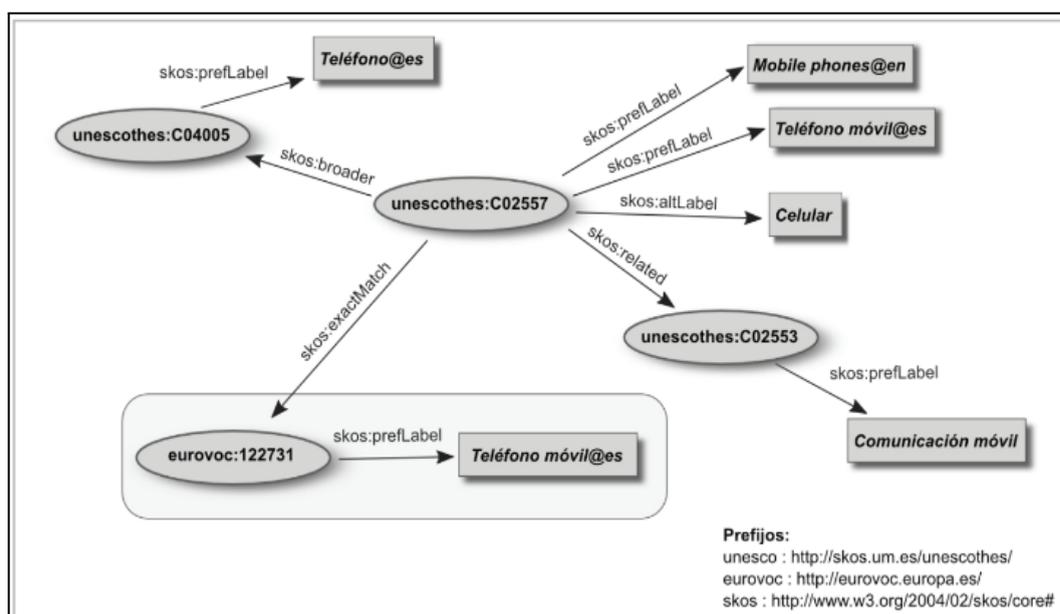
Na norma ISO 25964 “uma relação de equivalência é estabelecida entre dois termos em um tesouro representando o mesmo conceito na mesma linguagem natural (Martínez-González; Alvite-Díez, 2019c, p. [8], tradução livre)”. Este tipo de equivalência não deve ser confundido com as equivalências de mapeamento e alinhamento de conceitos SKOS inter-vocabulários controlados, tais como *skos:exactMatch* e *skos:approxMatch*. A notação *skos:closeMatch* relaciona conceitos similares para uso intercambiável em aplicações específicas, enquanto *skos:exactMatch* relaciona conceitos com alto grau de correspondência, tornando-os intercambiáveis em várias aplicações. (Miles; Bechhofer, 2011)

Normas anteriores à ISO 25964 adotavam uma abordagem léxica para tesauros. Essa abordagem envolvia definir uma série de termos descritores e não descritores, estabelecendo relações semânticas terminológicas. No entanto, essa abordagem apresenta problemas em termos de manutenção dos tesauros e processos de indexação. Com o SKOS, os documentos são associados a conceitos do tesouro, em vez de apenas termos. Isso significa que as mudanças nos termos descritores e não descritores não afetam a indexação no SKOS, pois apenas

envolvem modificações nos rótulos atribuídos aos conceitos. Deste modo, nos tesouros terminológicos, mudanças na estrutura dos termos descritores e não descritores requerem atualizações nos índices que relacionam os termos aos documentos, bem como nas relações estabelecidas com termos de outros tesouros. O uso do SKOS facilita a representação de tesouros multilíngues, em que até mesmo termos equivalentes em diferentes idiomas são representados como rótulos associados ao mesmo conceito. Com o SKOS, uma alteração na rotulagem dos conceitos não afeta a estrutura das relações semânticas (International Organization for Standardization, 2011; Isaac; Summers, 2012; ISO TC46/SC9/WG8; Isaac, 2012; Miles; Bechhofer, 2011; Pastor-Sánchez, 2013a).

A Figura 6 mostra um exemplo resumido de representação de um conceito do Tesouro da UNESCO com o SKOS e sua conexão com um conceito do Tesouro Eurovoc. Nela, é possível observar como termos equivalentes em diferentes idiomas são representados como etiquetas associadas ao mesmo conceito.

**Figura 6** - Correspondência entre o Tesouro da Unesco e o Eurovoc.



Fonte: Pastor-Sánchez (2013a)

Na Figura 6, os conceitos equivalentes são indicados por *unescothes:C02557* no Tesouro da Unesco e como *eurovoc:122731* no Tesouro Eurovoc. Os conceitos

são equivalentes por ambos representam exatamente (*skos:exactMatch*) uma mesma ideia, representando a ideia/conceito de “celular”.

É deste ponto de vista que a norma ISO 25964 trata as relações de equivalência em tesauros monolíngues separadamente do tratamento das relações de equivalência entre dois idiomas. Visto que o fato de as relações de equivalência ocorrerem a nível lexical (entre termos), ainda assim, os conceitos que esses termos representam precisam ser equivalentes. Deste modo, a ISO 25964 permite equivalência entre dois termos de um mesmo conceito no mesmo idioma em um tesouro mono ou multilíngue. As notações de equivalência são USE (para o termo preferido) e UF (para o termo não preferido). Essas notações são mantidas para compatibilidade com a ISO 2788, representando o mesmo significado que já designavam pelas relações de rótulo preferido/não preferido de cada conceito com seus rótulos (International Organization for Standardization, 2011).

As relações de equivalência estabelecidas pelo modelo de dados SKOS (Quadro 6, subseção 4.1.3) são utilizadas inter-SOC e, por conseguinte, podem ser usadas inter-tesauros. As relações de equivalência intra-SOC em SKOS são expressas pelas propriedades *skos:prefLabel* (rótulo preferido), *skos:altLabel* (rótulo não preferido, alternativo), *skosxl:prefLabel* (objeto rótulo preferido) e *skosxl:altLabel* (objeto rótulo não preferido, alternativo). Sendo as duas primeiras de “equivalências” entre termos e conceitos (para rotular um conceito), e as duas últimas permitem expressar equivalência de nível lexical entre termos preferenciais e não preferenciais de um mesmo conceito.

## 5.6 Equivalência composta

A parte dois da norma ISO-25964 lida com interoperabilidade de tesauros com outros vocabulários controlados, em que equivalências são definidas entre conceitos de diferentes vocabulários (Balkan; Bell, 2014), conforme exemplifica Pastor-Sánchez (2016): a) Equivalência exata: mesmas ideias e escopos idênticos. Exemplo: "Doença da vaca louca" e "Encefalopatia espongiforme bovina"; b) Equivalência inexata: escopos sobrepostos ou pequena diferença de conotação. Exemplo: "assentos" e "cadeiras"; c) Equivalência parcial: significado próximo, com um conceito mais abrangente. Exemplo: "Aeronave" e "Aviões"; d) Mapeamentos

hierárquicos: um conceito é mais abrangente que o outro. Exemplo: "Países Baixos" e "Holanda"; e f) Mapeamentos associativos: relação semântica entre conceitos não equivalentes. Exemplo: "Fotografias" e "Fotógrafos".

De acordo com a norma ISO-25964, uma relação de equivalência pode ocorrer entre conceitos de tesouros diferentes (equivalência simples) ou por combinação de conceitos em outro tesouro (equivalência composta). A equivalência composta pode ocorrer por meio da

[...] interseção ou da união do significado dos conceitos do tesouro de destino. As relações de equivalência não são explicitamente refletidas no modelo de dados da norma, mas podem ser derivadas a partir de relações hierárquicas e associativas (Pastor-Sánchez, 2013a, p. 191, tradução livre).

De acordo com a norma ISO 25964-2,

a equivalência pode ser "simples" (quando os dois tesouros contêm conceitos idênticos em escopo) ou "composta" (quando um conceito representado em um vocabulário com apenas um termo preferencial pode ser representado em outro vocabulário por uma combinação de dois ou mais conceitos/termos). Mapeamentos de equivalência simples também podem ser exatos ou inexatos (Balkan; Bell, 2014, p. 19, tradução livre).

A equivalência simples exata ocorre quando "os conceitos podem ser usados de forma intercambiável em todas as aplicações que podem ser imaginadas para o mapeamento" (International Organization for Standardization, 2013, p. 26, tradução livre). Em oposição, a equivalência simples inexata ocorre quando os conceitos são equivalentes em alguns contextos, mas não em outros, ou quando os conceitos têm escopos sobrepostos ou pequenas diferenças de conotação (Balkan; Bell, 2014).

Nesta perspectiva, o uso do SKOS ganhou ampla experiência devido ao seu uso na interconexão de vocabulários. O SKOS inclui propriedades para definir diversos tipos de equivalências, como exatas, próximas, hierárquicas e associativas. No entanto, essas propriedades permitem apenas estabelecer relações de equivalência simples, não compostas (Clarke; Zeng, 2012; Pastor-Sánchez, 2016).

A ISO 25964-2:2013 introduz a equivalência composta, tipo de relação que não era orientada a partir da norma ISO 2788:1986. Uma equivalência composta ocorre quando "um conceito complexo que é incluído em um vocabulário com apenas um termo preferido pode ser representado em outro vocabulário por meio de

uma combinação de dois ou mais conceitos/termos” (International Organization for Standardization, 2013).

Martínez-Gonzalez e Alvite-Díez (2019c) e Pastor-Sánchez (2016) argumentam que a definição e manutenção do nível lexical do tesauro com a norma ISO-25964 pode ser bastante complexa devido às relações de equivalência entre termos – *USE* (para termo preferencial; *skosxl:prefLabel* no SKOS-XL) /*USE+* e *UF* (para termo não preferencial; *skosxl:altLabel* no SKOS-XL) /*UF+*) – e as relações de equivalência entre termos e conceitos (*isNonPreferredlabelFor*, *isPreferredlabelFor*). Essas relações permitem a equivalência entre termos não preferenciais compostos (*SplitNonPreferredTerm*) e vários termos preferenciais (*PreferredTerm*) usando uma relação de equivalência composta (*CompoundEquivalence*) (International Organization for Standardization, 2011, 2013). O SKOS não possui elementos para definir essas equivalências, mesmo com a extensão SKOS-XL (SKOS *eXtension for Labels*) (Alvite-Díez; Martínez-González, 2021; Martínez-González; Alvite-Díez, 2019b;).

Para exemplificar a complexidade da equivalência composta Will (2012) escreve sobre o conceito composto “mineração de carvão” não existir em um tesauro, mas poder ser expresso

como uma combinação de dois ou mais conceitos mais simples que estão lá. Este caso é mostrado simbolicamente como:

*mineração de carvão*

*USE+ carvão*

*USE+ mineração*

com recíprocos como “*carvão UF+ mineração de carvão*” (Will, 2012, p. 49)

Um conceito composto não possui notações, atributos ou notas de escopo na norma ISO 25964, sendo interpretado com base nos escopos dos conceitos individuais. Num tesauro utilizado para indexação pós-coordenada, o indexador pode associar os termos “carvão” e “mineração” a um documento sem expressar uma relação entre eles, combinando esses termos a partir do operador booleano *AND*. Nesse sentido, a norma ISO 25964 justifica a ausência de elementos para representar um conceito composto:

Outro detalhe diz respeito ao “*USE...+*”. Nas visualizações para leitores humanos, sempre há um espaço entre “*USE*” e o termo intermediário (por exemplo, “*mineração de carvão USE carvão +*”).

mineração"). No entanto, se a marcação for usada nos bastidores por computadores, ela precisa ser comprimida, no estilo "mineração de carvão USE+ carvão"; "mineração de carvão USE+ mineração". Por esse motivo, o espaço é omitido no diagrama [exemplo na página 108 da norma que apresenta notações como NT e BT, todas para relações hierárquicas] (International Organization for Standardization, 2011, p. 109, tradução livre).

O SKOS também não possui elementos para expressar conceitos compostos (Alvite-Díez; Martínez-González, 2021; Martínez-González; Alvite-Díez, 2019b). Contudo, estruturas como equivalência composta e conceitos compostos não são comuns em tesouros, “para os quais a simplicidade é um princípio de design” (Martínez-González; Alvite-Díez, 2019c, p. 83, tradução livre).

## 5.7 Relações hierárquicas

A ISO 25964 mantém esses símbolos tradicionais, BT e NT, porém essas notações devem ser entendidas como "conceito mais amplo" e "conceito mais restrito", respectivamente (Martínez-González; Alvite-Díez, 2019a,c). Nesse contexto, o quadro a seguir apresenta a comparação entre a referida norma e o modelo SKOS.

**Quadro 13** - Comparação de relações hierárquicas na ISO 25964 e SKOS.

Relacionamento na o ISO 25964	Função ISO	Propriedade SKOS	Função SKOS
BT/NT a) sem tipo b) tipo = “tipo de” c) tipo = “parte de” d) tipo = “instância de”	relacionar dois <i>ThesaurusConcept</i> , de forma ampla/específica	<i>skos:broader</i> / <i>skos:narrower</i>	indicar relacionamento geral / indicar relacionamento específico

Fonte: Elaborado pela autora

Conforme apresentado no Quadro 13, a norma ISO 25964 traz outra novidade acerca das relações hierárquicas em comparação às normas anteriores, ao passo que formaliza a tipologia de relacionamentos hierárquicos em três tipos: gênero-espécie (tipo de), parte-todo ou positivo (parte de), e instancial (instância de) (Martínez-Gonzalez; Alvite-Díez, 2019c; Will, 2012). Como respectivos exemplos, cita-se: animal BTG cachorro (G de *generic*/genérico); computador BTP teclado;

planeta BTI terra. Todavia, essas distinções não possuem equivalentes no modelo SKOS.

As relações hierárquicas, como a relação BT nos padrões do tesauro, são cruciais nos sistemas de organização do conhecimento. Essa relação conecta entidades específicas a entidades mais genéricas, representando sua posição na hierarquia conceitual. O modelo SKOS formalizou essa relação como a propriedade *skos:broader*.

Além disso, os tesouros podem utilizar outros tipos de relações hierárquicas mais especializadas, conforme exemplificado pelos vocabulários do Getty publicados em LOD, como o *Art and Architecture Thesaurus (AAT)*<sup>23</sup>, o *Union List of Artist Names (ULAN)*<sup>24</sup> e o *Thesaurus of Geographic Names (TGN)*<sup>25</sup> (Alexiev; Isaac; Lindenthal, 2016)

Na norma ISO 25964 é introduzida a distinção entre três tipos de relacionamentos hierárquicos<sup>26</sup>: genérico (BTG), partitivo (BTP) e instancial (BTI). No entanto, essa distinção ainda não é mapeada no SKOS (Alvite-Díez; Martínez-González, 2021; Martínez-Gonzalez; Alvite-Díez, 2019b). Nesse sentido, Alexiev, Isaac e Lindenthal (2016) propõem como solução definir propriedades estendidas (BTGE, BTPE e BTIE) (ver subseção 5.12.3).

Will (2012) aponta que a tipologias de relacionamentos hierárquicos introduzidas pela ISO são convenientes até mesmo para subdividir esses tipos ainda mais, de modo a estabelecer diferentes níveis para cada tipologia (transitividade), se necessário; mas isso precisa ser possível de modo que as tipologias e subdivisões sejam desconsideradas por sistemas que não as utilizam.

O primeiro nível de distinção é relevante para sistemas automatizados e compatibilidade com ontologias, permitindo identificar se um relacionamento é transitivo ou não. Relações genéricas/específicas preservam a transitividade, enquanto relações todo/parte podem não preservar. A norma recomenda o uso

---

<sup>23</sup> <https://www.getty.edu/research/tools/vocabularies/aat/>

<sup>24</sup> <https://www.getty.edu/research/tools/vocabularies/ulan/>

<sup>25</sup> <https://www.getty.edu/research/tools/vocabularies/tgn/>

<sup>26</sup> *Broader Term Generic (BTG); Broader Term Partitive (BTP); BTI Broader Term Instancial (BTI)*

específico de relações partitivas em disciplinas, localizações geográficas, sistemas do corpo e estruturas sociais hierárquicas. As localizações geográficas são um caso especial devido aos seus nomes próprios, que rotulam instâncias individuais. O relacionamento instancial é usado para mostrar que uma instância é membro de uma classe (Will, 2012).

O quadro a seguir exemplifica os tipos de relacionamentos hierárquicos orientados pela norma ISO 25964:

**Quadro 14** - Tipos de relação hierárquica.

Relação	Abrev.	Nome	Exemplo
Conceito genérico mais amplo	BTG	Relação espécie/gênero	A calcita é um tipo de mineral (AAT)
Conceito partitivo mais amplo	BTP	Relação parte/todo	A Toscana faz parte da Itália (TGN)
Conceito instancial mais amplo	BTI	Relação instância/tipo	Rembrandt é uma instância de (exemplo de) uma pessoa (ULAN)

Fonte: adaptado de Alexiev, Isaac e Lindenthal (2016)

Ao analisar o Quadro 14, recorda-se que a norma esclarece que embora seja mantido o “T” da abreviatura, em designação convencional a termo, a relação é tratada essencialmente entre conceitos.

Embora essas relações não sejam comuns em tesouros clássicos, o escopo dos tesouros (AAT e TGN) e lista de autoridade (ULAN) utilizados para exemplificação favorece a identificação dos exemplos dispostos no Quadro 14. O AAT é um tesouro que traz quantidade significativa de elementos, artefatos e respectivas espécies/tipos. O ULAN trará nomes de artistas, favorecendo a identificação de instâncias como exemplos específicos de pessoas. Enquanto o TGN trará nomes geográficos, propiciando a identificação de partes de lugares, porém esses lugares também são instâncias, exemplos de lugares. Esta constatação tende a gerar divergências e necessidades de soluções *ad hoc* para representação de relações hierárquicas do tipo BTI e BTP. Por isto que relações genéricas/específicas preservam a transitividade, enquanto relações todo/parte podem não preservar.

As relações hierárquicas podem ser combinadas em um tesauro para criar estruturas poli-hierárquicas, nas quais cada conceito pode ter mais de um conceito superior. Essas estruturas são definidas pela ISO 25964-1:2011 como um arranjo hierárquico de conceitos que permite múltiplos conceitos superiores para cada conceito. Embora não sejam comuns (Martínez-Gonzalez; Alvite-Díez, 2019c), essas estruturas podem ser utilizadas para ampliar a flexibilidade e a representatividade dos tesouros e esquemas de classificação.

A poli-hierarquia não possui uma notação específica na norma ISO 25964 ou no modelo SKOS, constituindo-se em uma estrutura mais complexa do que uma árvore hierárquica em um tesauro. Modelar e raciocinar sobre um poli-hierarquia pode ser mais prático e confiável quando a hierarquia resultante é uma árvore. O SKOS não restringe o uso de poli-hierarquia, mas como não possui meios específicos para representá-la e controlá-la, a poli-hierarquia é uma característica específica dos tesouros e esquemas de classificação que é controlada por soluções *ad hoc*, conforme finalidades de uso (Martínez-Gonzalez; Alvite-Díez, 2019b; Alvite-Díez; Martínez-González, 2021).

### **5.8 Principais termos (*topTerms*) e principais conceitos (*topConcepts*)**

Os principais conceitos (*topTerms*) e termos (*topConcepts*) são mais amplos do que qualquer outro conceito no domínio (grupo de conceitos) ao qual pertencem. Conforme a norma ISO 25964, cada conceito pode estar vinculado ao conceito no topo de qualquer hierarquia em que esteja inserido. Esses conceitos principais podem ser nomes de facetas (atividades, pessoas, lugares ou coisas), por exemplo, e essa ligação facilita a navegação, mostrando claramente em qual faceta um conceito se encontra. Além disso, pode ser usado para validação, considerando-se que os relacionamentos hierárquicos são válidos apenas se os dois conceitos estiverem na mesma faceta. Um conceito também pode ter um atributo booleano (verdadeiro/falso) para indicar se é um "conceito principal". Isso pode ser útil para criar uma lista de conceitos de alto nível a partir dos quais iniciar a navegação (Will, 2012). Os conceitos principais podem ser identificados navegando para cima na hierarquia até que nenhum conceito mais amplo seja encontrado.

O Quadro 15 compara relacionamentos para conceitos principais na norma ISO 25964 e no modelo de dados SKOS.

**Quadro 15** - Relacionamentos para conceitos principais na ISO 25964 e SKOS.

Relacionamento na ISO 25964	Função ISO	Propriedade SKOS	Função SKOS
<i>hasTopConcept/ isTopConceptOf</i>	relacionar <i>ThesaurusConcept</i> , e <i>ConceptGroup</i>	<i>skos:topConceptOf / skos:hasTopConcept</i>	relacionar um <i>skos:Concept</i> a um <i>skos:ConceptScheme</i>

Fonte: Elaborado pela autora

Tesauros que utilizam *skos:ConceptScheme* para representar microtesauros usam a propriedade *skos:hasTopConcept* para representar os principais conceitos dos microtesauros. Contudo, quando *skos:Collection* são usadas para representar microtesauros, não é possível utilizar *skos:hasTopConcept* (Martínez-Gonzalez; Alvite-Díez, 2019b,c). Esse impasse também é outro ponto em que há necessidade de solução *ad hoc* para explicitar um conceito principal no contexto de RI.

## 5.9 Relações associativas

A ISO 25964 oferece a possibilidade de comentar a natureza das relações associativas, tais como: causa/efeito, processo/produto ou pessoa/disciplina, exemplificando, acidente/lesão, controle de temperatura/termômetro ou bibliotecário/Biblioteconomia, respectivamente. Essa abordagem aproxima os tesauros das ontologias, onde a natureza de todos os relacionamentos é especificada (Martínez-Gonzalez; Alvite-Díez, 2019a,c; Will, 2012), mas possui notação similar em SKOS.

A seguir, tem-se o quadro comparativo das notações para relacionamentos associativos.

**Quadro 16** - Relacionamentos associativos em ISO 25964 e SKOS.

Relacionamento na ISO 25964	Função SKOS	Relacionamento no SKOS	Função SKOS
RT	relacionar	<i>skos:related</i>	designar relação

	<i>ThesaurusConcept a ThesaurusConcept</i>		associativa entre dois <i>skos:Concept</i>
--	--	--	---

Fonte: Elaborado pela autora

Não foi identificada a possibilidade de relações associativas entre microtesauros, coleções, grupos de conceitos ou domínios, tanto nas recomendações SKOS, quanto na norma ISO 25964.

## 5.10 Notas

Orientadas pela ISO 25964 e pelas recomendações SKOS, assim como nas normas anteriores, as notas não mudam a essência dos tesouros, por isto geralmente pouca discussão é dedicada às mesmas na literatura. (Martínez-González; Alvite-Díez, 2019a,c).

Will (2012) afirma que o modelo de dados ISO 25964 permite associar notas a conceitos e termos, incluindo notas personalizadas para necessidades específicas. Também incorpora atributos de outros padrões, como o esquema de metadados Dublin Core (ISO TC46/SC9/WG8; Isaac, 2012; Will, 2012).

Também é possível registrar um histórico de versões em um tesouro, mostrando as diferentes versões e suas distinções, além de poder indicar sua atualidade. Há a possibilidade de adicionar datas de criação e modificação a cada conceito e termo (International Organization for Standardization, 2011).

Os elementos de notas são comparados no Quadro 17 a seguir:

**Quadro 17** - Comparação entre elementos de notas da ISO 25964 e do SKOS.

<b>Elemento de anotação</b>	<b>ISO 25964</b>	<b>SKOS</b>
Notas gerais	<i>Note</i>	<i>skos:note</i>
Notas de escopo	<i>ScopeNote</i>	<i>skos:scopeNote</i>
Notas de histórico	<i>HistoryNote</i>	<i>skos:historyNote</i>
Notas de lançamento/publicação	<i>EditorialNote</i>	<i>skos:editorialNote</i>
Definições	<i>Definition</i>	<i>skos:definition</i>

Exemplos	-	<i>skos:example</i>
Notas de alteração	-	<i>skos:changeNote</i>
Notas personalizadas	<i>CustomNote</i>	-

Fonte: Pastor-Sánchez (2016).

Notas esclarecem o significado e aplicação de termos ou conceitos em relação a outros do mesmo tesouro. Os editores e desenvolvedores de tesouros podem adicionar ou não quantas notas forem necessárias (Alvite-Díez; Martínez-González, 2021; International Organization for Standardization, 2011; Martínez-González; Alvite-Díez, 2019b).

### 5.11 Arranjos

Arranjos (*Array*) servem para representar domínios e arranjos. Um *Array* da norma ISO 25964 é correspondente a uma *skos:Collection*, constituindo-se em uma coleção/conjunto de conceitos que possuem um “conceito pai” comum (International Organization for Standardization, 2011; ISO TC46/SC9/WG8; Isaac, 2012; Martínez-González; Alvite-Díez, 2019a,c) como, por exemplo, ferramenta: “veículo’: 'veículo agrícola, almofada de ar veículo, veículo de acampamento, veículo elétrico, veículo grande, veículo motorizado” (Martínez-González; Alvite-Díez, 2019c, p. [13], tradução livre). Estes conceitos derivados do “conceito pai” podem ser considerados “conceitos irmãos”. Logo, arranjos são compostos por conceitos irmãos. Nesse sentido, os arranjos são estruturas semelhantes aos SOC do tipo anéis de sinônimos.

Will (2012), no concernente a arranjos e rótulos de nós (*Node Labels*) previstos pela ISO 25964-1:2011, pontua que grupos de conceitos irmãos podem ser organizados em arranjos usando rótulos de nós. Esses rótulos são úteis para navegação, pesquisa e seleção de termos em exibições hierárquicas de tesouros. A ordem dos conceitos em um arranjo pode ser diferente da ordem alfabética dos termos preferidos. Os rótulos de nó não representam conceitos nem têm relações hierárquicas ou associativas com conceitos. Embora em alguns *softwares* de tesouros os rótulos de nó sejam tratados como termos preferenciais ou não preferenciais. No caso do exemplo de “conceito pai” e respectivos “conceitos filhos”,

um rótulo de nó seria “tipo de veículo” (característica comum a todos os “conceitos irmãos”).

O quadro a seguir mostra comparação entre relacionamentos instruídos na norma ISO 25964 e modelo SKOS para arranjos:

**Quadro 18** - Relacionamentos para arranjos na ISO 25964 e no modelo SKOS.

Relacionamento na ISO 25964	Função ISO	Propriedade no SKOS	Função SKOS
hasMemberConcept / isMemberOfArray	relacionar <i>ThesaurusConcept</i> e <i>ThesaurusArray</i>	<i>skos:Collection</i>	indicar um arranjo de conceitos

Fonte: Elaborado pela autora.

Em consonância com o quadro 18, Martínez-González e Alvite-Díez (2019b) evidenciam que há propostas para representar, no SKOS, os arranjos e domínios com *skos:Collection*.

## 5.12 Abordagens de aprimoramento, extensões e usos

Alguns dos textos analisados apresentam abordagens de aprimoramento tanto com relação às instruções da norma ISO 25964, quanto ao modelo de dados SKOS, de acordo com necessidades específicas de representação temática da informação via tesouros conceituais e suas respectivas áreas do conhecimento e necessidades de uso.

Considerar abordagens de aprimoramento, extensões e usos do modelo de dados SKOS e da norma ISO 25964, no contexto de suas similaridades, implica diretamente em abordagens e aplicações relacionadas à conversão, engenharia e reengenharia de tesouros. Deste modo, extensões SKOS e casos de uso da norma e do modelo são apresentados nesta subseção.

### 5.12.1 Mapeamento entre tesouros

Balkan e Bell (2014) apresentam resultados parciais de um projeto de vinculação de tesouros no contexto da Web Semântica e LOD. Nesse sentido, os

autores esclarecem que o Arquivo de Dados do Reino Unido<sup>27</sup> gerencia dois tesouros: o tesouro multilíngue *European Language Social Science Thesaurus* (ELSST)<sup>28</sup> e o tesouro monolíngue *Humanities and Social Science Electronic Thesaurus* (HASSET)<sup>29</sup>, do qual o ELSST é derivado (Balkan; Bell, 2014).

A partir de parceria com o *Consortium of European Social Science Data Archives* (CESSDA)<sup>30</sup>, através do projeto CESSDA-ELSST e por meio do financiamento do *Economic and Social Research Council* (ESRC) do Reino Unido, o Arquivo de Dados do Reino Unido atualizou ambos os tesouros, que passaram a ser baseados em conceitos e não em termos. O projeto se fundamentou no cumprimento da ISO 25964 (parte 1 e 2) e no modelo de dados SKOS, juntamente com aplicações específicas de gerenciamento, criando conexões entre os tesouros. Estendendo a ISO 25964, o Arquivo desenvolveu uma forma de mapear seus tesouros de ciências sociais para facilitar a recuperação de dados transnacionais arquivísticos. Também estendeu o modelo SKOS e desenvolveu uma nova aplicação para o gerenciamento de tesouros que combina visualização de termos com estruturas de árvore. Os dois tesouros, então, passaram a compartilhar um conjunto de conceitos centrais claramente definidos, mesmo com algumas divergências. O novo sistema de gerenciamento de tesouros permite que os termos sejam promovidos para o conjunto central quando apresentarem equivalência parcial ou exata, ou sejam rebaixados para 'não centrais'. O sistema também permite que equivalentes linguísticos autorizados sejam adicionados aos termos do ELSST. Nesse contexto, o sistema de gerenciamento de tesouros proposto pelo projeto também fornece sugestões agrupadas para alterações nos termos ou estrutura do tesouro, vinculadas à estrutura de árvore (Balkan; Bell, 2014).

Balkan e Bell (2014) mostram como a existência da norma ISO 25964-2:2013 foi fundamental para estabelecer o mapeamento entre dois tesouros do Arquivo de Dados do Reino Unido. No entanto, ressaltam que as definições de equivalência que

---

<sup>27</sup> <http://www.data-archive.ac.uk>

<sup>28</sup> <http://elsst.ukdataservice.ac.uk>

<sup>29</sup> <http://hasset.ukdataservice.ac.uk>

<sup>30</sup> O CESSDA é uma organização que promove a aquisição, arquivamento e distribuição de dados eletrônicos para o ensino e pesquisa em ciências sociais na Europa; disponível em: <https://www.cessda.eu/>

elaboraram para ampliar o SKOS são mais rigorosas devido às soluções *ad hoc* elaboradas para as relações de equivalência.

O trabalho do projeto CESSDA-ELSST exemplifica como as similaridades entre a norma ISO 25964 e o modelo de dados SKOS podem ser aplicadas num contexto real. Balkan e Bell (2014, p. 17, tradução livre) afirmam que

[...] os principais objetivos do projeto CESSDA-ELSST eram, portanto, converter os tesauros de um modelo baseado em termos para um modelo baseado em conceitos, seguindo a ISO 25964, e disponibilizá-los como Dados Conectados baseados em SKOS.

Para Bandyopadhyay e Mukhopadhyay (2015), em conformidade com a norma ISO 25964-2, o mapeamento é um processo de estabelecimento de relações entre os conceitos de um vocabulário com outro para trocar e usar informações entre dois ou mais sistemas.

No contexto de construção de tesauros multilíngues o mapeamento é crucial para estabelecimento de interoperabilidade entre tesauros de diferentes idiomas, em que as equivalências serão estabelecidas entre termos que representam os conceitos em diferentes línguas.

Ao abordarem a importância da conversão de tesauros clássicos para tesauros conceituais interoperáveis, no âmbito do processo de projetar um tesouro multilíngue, Bandyopadhyay e Mukhopadhyay (2015) apontam que tesauros projetados conforme a norma anterior à ISO 25964 foram utilizados para armazenar e recuperar informações de mídia impressa. No entanto, atualmente, é necessário projetar tesauros que possam recuperar dados de bancos de dados eletrônicos e se integrar com outros sistemas, como os OPACs de bibliotecas (Bandyopadhyay; Mukhopadhyay, 2015). Tesauros legíveis por máquina e interoperáveis

[...] têm algumas vantagens sobre os impressos [ou em PDF], pois são facilmente editáveis, portáteis, gerenciáveis, compartilháveis, pesquisáveis, seguros no manuseio, a saída pode ser impressa em vários formatos [etc.] (Bandyopadhyay; Mukhopadhyay, 2015, p. 14, tradução livre).

A interoperabilidade requer mapeamento entre vocabulários para apresentação de dados em um formato padrão, além de uma plataforma de computador comum usando os mesmos padrões e protocolos. As demandas por interoperabilidade também podem levar ao aprimoramento das funcionalidades das

ferramentas de gerenciamento de SOC, ao mesclar vários vocabulários ou usar parte de um para estender outro (Bandyopadhyay; Mukhopadhyay, 2015).

Dentre as ferramentas que se destacaram durante a análise dos resultados desta pesquisa, cita-se: Tematres, SKOSEd, PoolParty, Skosmos, VocBench, SKOS-play e Opentheso.

O Tematres<sup>31</sup> condiciona interoperabilidade exportando seus dados para vários formatos padrão, sendo uma ferramenta útil para gerenciar tesouros, pois atende à maioria das qualidades essenciais da ISO 25964 (1 e 2). É compatível com SKOS e de código aberto, o que o torna uma opção eficaz para construir tesouros digitais multilíngues e vinculados interoperavelmente (Bandyopadhyay; Mukhopadhyay, 2015).

Martínez-González e Alvite-Díez (2019) propuseram um *framework* metodológico para avaliar a qualidade de ferramentas de gerenciamento de tesouros, de modo a comparar a norma ISO 25964 e o modelo de dados SKOS e, assim, fornecer uma prova de conceito para o *framework* proposto. As ferramentas analisadas são o SKOSEd e o PoolParty. O SKOSEd é uma ferramenta criada para editar e gerenciar SOC de diversos tipos em SKOS. O PoolParty é uma ferramenta criada especificamente para tesouros.

Além dessas ferramentas, o Skosmos e o VocBench foram mencionados em algumas publicações. O primeiro se presta como ferramenta de publicação web, busca e navegação de SOC. Enquanto o segundo serve para edição de SOC. O VocBench é um exemplo de aplicativo Web de código aberto para edição de tesouros modelados em SKOS e SKOS-XL (Alvite-Díez; Martínez-González, 2021; Berti Junior *et al.*, 2017; Caracciolo *et al.*, 2012; Martínez-González, Alvite-Díez, 2019b,c; Stellato, 2015a,b).

Zeng e Mayr (2018) apresentam diversas ferramentas para várias finalidades no contexto LOD. O SKOS-play<sup>32</sup>, por exemplo, é uma ferramenta gratuita que permite representar e visualizar tesouros, taxonomias ou listas controladas expressas em SKOS. É especialmente útil para usuários que não estão

---

<sup>31</sup> <https://sourceforge.net/projects/tematres/>

<sup>32</sup> <http://labs.sparna.fr/skos-play/>

familiarizados com linguagens de marcação, pois oferece a capacidade de converter planilhas do Excel em arquivos SKOS, além de fornecer recursos de visualização.

Saliba *et al.* (2021) mencionam o *software* Opentheso, empregado para a criação de tesouros de código aberto. Essa ferramenta está em conformidade com as duas partes da norma ISO 25964 e suporta SKOS. O Opentheso permite a publicação *online* de tesouros por meio da integração com outras plataformas.

### 5.12.2 Extensões SKOS

As similaridades e as diferenças entre a norma ISO 25964 e o modelo de dados SKOS não são mera coincidência. Will (2012) afirma que o modelo de dados SKOS é amplamente compatível com a norma ISO 25964, mas ainda não implementa todos os recursos desta. Na época de lançamento da norma, já havia diversas discussões sobre possíveis extensões para SKOS com o intuito de cobrir recursos previstos pela norma ISO 25964.

Ao longo dos textos selecionados para esta pesquisa, identificaram-se algumas extensões SKOS, como a extensão FinnONTO SKOS, que define propriedades *broadGeneric*, *broadPartitive*. Deste modo, nesta subseção, a atenção será dada às principais extensões identificadas no *corpus* de análise.

O *SKOS eXtension for Labels* (SKOS-XL), o UNESKOS *Vocabulary Specification* e o ISO-THES estão entre as tecnologias para aprimoramento do modelo SKOS. O primeiro surge como modelo de extensão para mitigar as limitações originais do SKOS. Enquanto o segundo e o terceiro surgem como modelos de extensão para complementar o SKOS de acordo com as normas ISO 25964:2011 e ISO 25964:2013 (Barbosa; Viera, 2022; Pastor-Sánchez, 2016;). O segundo e o terceiro modelos de extensão são desenvolvidos pelo grupo de trabalho ligado ao vocabulário da UNESCO (Pastor-Sánchez, 2016).

Quando os “modelos de dados para SKOS e ISO 25964-1 estavam em desenvolvimento, as equipes responsáveis por eles mantiveram uma boa comunicação” (Clarke; Zeng, 2012, p. 26, tradução livre), resultando na ampla compatibilidade entre ambos os padrões principalmente quando a extensão SKOS-XL é levada em consideração. Em alguns casos, as extensões SKOS

representam semânticas semelhantes com propriedades diferentes. Em outros casos, a semântica é realmente específica de cada tesouro (Martínez-González; Alvite-Díez, 2019b). As especificidades surgem conforme os contextos e finalidades de uso dos tesouros representados.

Após o lançamento do modelo de dados SKOS *Core Vocabulary Specification*, em 2005, a sua especificação foi substituída, em 2009, pela recomendação oficial do W3C SKOS *Reference*. Nesta última versão, o modelo SKOS *Core*

[...] é complementado em seu Apêndice por um *eXtension for Labels* (SKOS-XL)<sup>33</sup>. Além de tudo o que é transmitido pelo SKOS *Core* para relacionamentos entre conceitos, a extensão fornece suporte adicional para identificar, descrever e vincular entidades lexicais [termos/rótulos]” (Clarke; Zeng, 2012, p. 23, tradução livre).

Berti Junior *et al.* (2017, p. 389) atestam que o “*Simple Knowledge Organization System eXtension for Labels* (SKOS-XL) [...] provê suporte adicional ao SKOS para descrição e ligação de entidades lexicais”. Neste sentido, a equipe responsável pelo desenvolvimento da norma ISO 25964, aponta que:

Quando o modelo de dados básico do SKOS não possui uma construção que corresponda a um recurso do modelo do ISO 25964, o modelo SKOS-XL é utilizado, complementado por propostas adicionais [como outras extensões e ontologias] quando necessário. Foi tomado cuidado para evitar qualquer incompatibilidade com outro projeto em andamento, visando alinhar o SKOS com o MADS [*Metadata Authority Description Schema*<sup>34</sup>] (National Information Standards Organization, 2023, p. [1], tradução livre)

Como exemplo específico de suprimento de limitação do SKOS *Core* para representação de tesouros, a partir do modelo de extensão SKOS-XL, está a classe *skosxl:Label*, que surge a partir da constatação de que as propriedades de rotulagem *skos:prefLabel*; *skos:altLabel* e *skos:hiddenLabel* são restritas a literais (valores) RDF e o SKOS *Core* não permite que literais RDF sejam declaradas como sujeitos (recursos conceituais) ou objetos, dado que não permite a relação entre etiquetas (rótulos/ termos). Contudo, alguns aplicativos requerem a criação de relações explícitas entre etiquetas atribuídas para um mesmo conceito. Por exemplo, ao considerar a relação entre uma etiqueta (literal) preferida para um conceito

<sup>33</sup> <http://www.w3.org/TR/skos-reference/skos-xl.html>

<sup>34</sup> US Library of Congress. MADS (Metadata Authority Description Schema). Washington, DC: Library of Congress. Disponível em: <http://www.loc.gov/standards/mads/>

"Corporation" e sua abreviatura "Corp.", cunhada como uma etiqueta alternativa ou uma relação de tradução entre duas etiquetas para um mesmo conceito representado por termos de idiomas diferentes: "Cow"@en e "Vache"@fr (Isaac; Summers, 2009); como solução para este problema, o *skosxl:Label* denomina uma classe do SKOS-XL onde as etiquetas (literais) também poderão ser entendidas como sujeitos de uma declaração RDF (Isaac; Summers, 2009).

O quadro a seguir ilustra os elementos SKOS-XL que ampliam o SKOS.

**Quadro 19** - Elementos da extensão SKOS-XL.

Elementos	Descrição
<i>skosxl:Label</i>	Classe especial de entidades lexicais (etiquetas)
<i>skosxl:literalForm</i>	Forma literal de uma entidade lexical (etiqueta)
<i>skosxl:prefLabel</i>	Etiquetas preferidas
<i>skosxl:altLabel</i>	Etiquetas alternativas
<i>skosxl:hiddenLabel</i>	Etiqueta oculta
<i>skosxl:labelRelation</i>	Serve como um ponto de extensão, que pode ser refinado para cenários de rotulagem mais específicos.

Fonte: Adaptado de Isaac e Summers (2009) e Barbosa e Viera (2022).

Atividades de criação de esquemas unificados com rótulos multilíngues usando SKOS-XL têm tido sucesso em várias partes do mundo, conforme relatado em vários casos (Stellato *et al.*, 2015a,b; Zeng; Mayr, 2018).

De acordo com Caracciolo (2012), o tesouro AGROVOC adotou o modelo SKOS-XL para atender aos requisitos de modelagem de um tesouro multilíngue e linguisticamente detalhado, já que a extensão SKOS-XL permite que os rótulos tenham relacionamento entre si e possam ser enriquecidos com propriedades próprias.

A extensão SKOS-XL permite uma modelagem linguística refinada, incluindo relações lexicais entre rótulos, além de detalhes mais específicos nas notas editoriais, como revisão de conceitos e rótulos em cada idioma (Caracciolo, 2012).

Deste modo, junto a outros mecanismos, interoperabilidade com outros SOC e soluções *ad hoc* que condicionam a disponibilidade do AGROVOC, Caracciolo (2012) aponta que o tesauro é usado em diversas situações (com maior interesse por pesquisadores acadêmicos e profissionais de bibliotecas) como, por exemplo: integração de dados estatísticos, textuais e geográficos da FAO; interlíngua para combinar diferentes conjuntos de dados; apoio a serviços web de tradução, bibliotecas, terminologia, tradução, indexação e gestão do conhecimento.

Após o lançamento da ISO 25964, que estabeleceu novos requisitos para o SKOS no tangente à representação de tesouros, um grupo de trabalho (equipe do projeto UNESKOS) começou a desenvolver a ontologia ISO-THES<sup>35</sup>. Essa ontologia complementa o SKOS e incorpora as propriedades da norma ISO 25964, permitindo a representação de tesouros adequados a diversos contextos de uso (Alvite-Díez; Martínez-González, 2021; Martínez-González; Alvite-Díez, 2019b; Pastor-Sánchez, 2016)

Pastor-Sánchez (2013a) esclarece que um dos aspectos mais complexos da norma ISO 25964 e da extensão ISO-THES, para o modelo de dados SKOS, “é a equivalência entre termos, principalmente a equivalência composta, embora seu uso seja desnecessário em muitos tesouros” (Pastor-Sánchez, 2013a, p. 192).

A extensão ISO-THES é uma solução altamente eficaz para alcançar a interoperabilidade semântica entre vocabulários baseados em SKOS e tesouros que seguem a norma ISO 25964. Por meio do ISO-THES, são estabelecidas conexões entre o esquema XML e o esquema RDF, permitindo que a maioria dos elementos da norma tenha equivalentes em SKOS, sem a necessidade de criar outro modelo de dados. Isso oferece aos editores de tesouros desenvolvidos de acordo com a norma ISO 25964 um caminho quase direto para publicar seus dados como um conjunto de dados RDF (ISO TC46/SC9/WG8; Isaac, 2012; Pastor-Sánchez, 2013a).

---

<sup>35</sup> Martínez-González e Alvite-Díez (2019c) afirmam que a documentação da extensão ISO-THES não está mais disponível. No entanto, a seguinte fonte apresenta algumas informações: <https://lov.linkeddata.es/dataset/lov/vocabs/iso-thes>

A extensão ISO-THES apresenta novas classes que ajudam a expressar relações terminológicas de equivalência composta, que ocorrem quando um conceito de um tesouro é representado pela combinação de vários conceitos (como microtesauros e outros esquemas de conceitos). A extensão ISO-THES contempla a complexidade de representar a relação entre termos não preferidos e dois termos preferidos (Pastor-Sánchez, 2016).

Dentre algumas opções de solução, Pastor-Sánchez (2016) aponta que para aplicar a extensão ISO-THES no Tesouro da UNESCO foram estabelecidos diversos critérios de modelagem que consideraram os seguintes elementos ISO-THES: *iso-thes:ConceptGroup*, *iso-thes:microThesaurusOf*, *iso-thes:supergroup* e *iso-thes:subgroup*. O autor evidencia, ainda, que a modelagem do tesouro com a extensão ISO-THES permitiu a adequação do SKOS com relação à norma ISO 25964, no âmbito das necessidades de representação e uso do Tesouro da UNESCO. O quadro a seguir apresenta elementos ISO-THES que estendem o SKOS.

**Quadro 20** - Exemplos de elementos da extensão ISO-THES.

Elementos	Descrição
<i>iso-thes:CompoundEquivalence</i>	Para representar a relação de equivalência composta entre um termo preferido e um termo não preferido.
<i>iso-thes:ConceptGroup</i>	Para representar um grupo de conceitos que constituem um subconjunto do tesouro.
<i>iso-thes:microThesaurusof</i>	Para vincular um termo a um grupo de conceitos do tesouro, reconhecido como um microtesouro.
<i>iso-thes:supergroup</i>	Para vincular um termo a um grupo superior de conceitos do microtesouro.
<i>iso-thes:subgroup</i>	Para vincular um termo a um grupo subordinado de conceitos do microtesouro.
<i>iso-thes:ThesaurusArray</i>	Para vincular grupos de conceitos irmãos.

Fonte: Barbosa e Viera (2022).

Propostas de extensão como a ISO-THES são interessantes, mas não são suportadas pela maioria das ferramentas de gestão de tesouro, o que implica que a

responsabilidade de transferir essas propostas adequadamente para seus tesouros continua a recair sobre os editores dos tesouros. Seria necessário que extensões como o ISO-THES evoluíssem para o *status* de Recomendação do W3C, permitindo que os desenvolvedores de *software* fornecessem suporte ao ISO-THES por padrão, de modo a facilitar a tarefa dos profissionais responsáveis pela edição dos tesouros (Martínez-González; Alvite-Díez, 2019c)

Por sua vez, o UNESKOS *Vocabulary Specification*<sup>36</sup> traz novas propriedades para relacionar conceitos a coleções de conceito ou esquemas de conceito de diferentes formas (Barbosa; Viera, 2022; Pastor-Sánchez, 2016).

Barbosa e Viera (2022, p. 236) esclarecem que

Um dos gatilhos para o desenvolvimento do UNESKOS foi a dificuldade em representar áreas de conhecimento que precisavam ser estruturadas em grupos e subgrupos de conceitos, reconhecidos como microtesouros. O desenvolvimento do UNESKOS se tornou necessário uma vez que o SKOS não fornece nenhum elemento nativo para uma representação direta dessas estruturas. Por esse motivo, o UNESKOS busca complementar o SKOS cumprindo as especificações da ISO-25964, na medida em que facilita a representação de qualquer tipo de tesouro, estruturando-os segundo grupos, subgrupos e supergrupos de classes de um domínio específico.

Dentre algumas opções de solução, Pastor-Sánchez (2016) indicia que para aplicar a extensão UNESKOS no Tesouro da UNESCO foram estabelecidos diversos critérios de modelagem que consideraram os seguintes elementos ISO-THES: *uneskos:memberOf*, *uneskos:mainConcept* e *uneskos:mainConceptOf*. Segue alguns exemplos de elementos do UNESKOS:

**Quadro 21** - Exemplo de elementos do UNESKOS *Vocabulary Specification*.

Elementos	Descrição
<i>uneskos:contains</i>	propriedade inversa do <i>skos:inScheme</i> que permite definir os elementos que estão contidos em um esquema de conceito.
<i>uneskos:hasMainConcept</i>	propriedade que permite definir certos conceitos como pontos de acesso às coleções de conceito.

<sup>36</sup> <https://skos.um.es/TR/uneskos/>

<i>uneskos:hasMicroThesaurus</i>	propriedade proposta como um mecanismo para descobrir o(s) grupo(s) de conceito(s) que conformam um esquema de conceito. Define o caminho para o grupo de conceitos, ao qual, o conceito pertence.
<i>uneskos:mainConceptOf</i>	Conecta um conceito a uma coleção.
<i>uneskos:memberOf</i>	propriedade inversa do skos:member para identificar a(s) Coleção (ões) às quais um conceito pertence. Esta propriedade permite definir links dos conceitos para suas coleções.

Fonte: Barbosa e Viera (2022).

A extensão UNESKOS foi mencionada em apenas duas publicações do *corpus*. No entanto, o que determina a utilização de uma extensão, bem como o desenvolvimento de uma nova extensão SKOS, além das necessidades de atendimento à norma ISO 25964, são as especificidades de uso.

### 5.12.3 Relações hierárquicas e transitividade

A norma ISO 25964 formalizou as relações hierárquicas transitivas nos tesouros conceituais, porém sem instruções claras de como estabelecê-las na prática (Alexiev; Isaac; Lindenthal, 2016). Essas relações favorecem o refinamento semântico em tesouros, propiciando o encontro acidental de informações a partir de inferências lógicas e conceituais válidas.

Em muitos casos, as relações hierárquicas podem ser transitivas, como quando um conceito é mais amplo que outro, que por sua vez é mais amplo que um terceiro. No entanto, em certas situações, especialmente em esquemas diferentes de tesouros, hierarquias mais complexas podem não ser adequadas. Por exemplo, se "veículos" for considerado mais amplo que "carros", e "carros" for considerado mais amplo que "rodas", pode ser problemático inferir automaticamente que "rodas" é mais específico do que "veículos". Nesse sentido, representar hierarquias transitivas instituídas pela ISO 25964 exige que seja mantida compatibilidade semântica com a propriedade *skos:broader* do SKOS. Deste modo, para Martínez-González e Alvite-Díez (2019b), em SKOS a transitividade se aplica quando as relações hierárquicas usadas para o raciocínio são todas genéricas (tipo de), mas não se misturam com relações partitivas (parte de).

Alexiev, Isaac e Lindenthal (2016) afirmam que poucos conjuntos de dados usam as propriedades de relação hierárquica transitiva da ISO 25964. Uma razão é a relativa novidade da norma, mas existem outras razões, tais como: axiomatização inadequada para composição de propriedade, gerando inferências lógicas errôneas; e relacionamentos mais amplos/estritos ausentes para não-conceitos.

Conforme tratado na subseção 5.7, os tesauros AAT e TGN, bem como a lista de autoridade ULAN possibilitam a identificação diversificada de exemplos dos três tipos de relação hierárquica instituídas pela ISO 25964: BTG, BTP e BTI, respectivamente. Diante desta constatação, Alexiev, Isaac e Lindenthal (2016) utilizam exemplos dos referidos SOC para analisar e definir propriedades estendidas (BTGE, BTPE e BTIE)<sup>37</sup> que contemplem aspectos de transitividade. Assim, como exemplo, uma propriedade estendida para BTG seria BTGE, onde “Se X é um tipo de Y e Y é um tipo de Z, então X é um tipo de Z” - Exemplo (AAT): *assadeiras* BTG *Bakeware* BTG *(vasilhas para cozinhar alimentos)* implica *assadeiras* BTGE *(vasilhas para cozinhar alimentos)*.

Nesse exemplo, percebe-se que “vasilhas para cozinhar alimentos” se trata de um termo guia que, por sua vez, indica a “faceta de tipos de conceito” (um não conceito). A abordagem de “faceta de tipos de conceitos”, bem como de análises lógicas, foi fundamental para fundamentar a proposta de propriedades estendidas de Alexiev, Isaac e Lindenthal (2016) para expressar relações hierárquicas transitivas formalizadas pela ISO 25964.

#### 5.12.4 Cenário geral de utilização

O cenário geral de utilização de tesauros representados de acordo com o modelo de dados SKOS, a norma ISO 25964 e outros modelos e diretrizes desenvolvidos de acordo com o contexto de uso, está diretamente relacionado às funções de interoperabilidade e mapeamento entre SOC de todos os tipos possíveis. Nesta perspectiva, Lara (2013), ao considerar o desenvolvimento da segunda parte da norma ISO 25964, já defendia que, para a realização de mapeamento, ou a

---

<sup>37</sup> *Broader Term Generic Extended* (BTGE); *Broader Term Partitive Extended* (BTPE); *Broader Term Instantial Extended* (BTIE)

criação de relacionamento entre dois ou mais tesouros ou outros tipos de esquemas de organização do conhecimento, seria necessário que modelos para cada SOC fossem analisados lado a lado, de modo a identificar os relacionamentos entre conceitos de cada um dos SOC.

Zeng e Mayr (2018) apresentam um panorama sobre o que o movimento LOD KOS trouxe para as comunidades de usuários de vocabulário controlados. O conjunto de LOD KOS (ver subseção 4.1) apresentados pelos autores é amplo e diverso. No entanto, os tesouros são SOC em destaque. Nessa perspectiva, foram detectados três contextos de utilização: 1) produção de conjuntos de dados LOD, em que os dados dos vocabulários LOD KOS se tornam LOD com classificação de 4 a 5 estrelas<sup>38</sup>; 2) produção de vocabulários controlados que favorecem o desenvolvimento e enriquecimento de SOC, servindo de base para novas descobertas e resultados inovadores; 3) pesquisas feitas por usuários finais (aqueles que não estão envolvidos com a produção de vocabulários ou conjuntos de dados), em que os produtos LOD KOS podem ser tornar base de conhecimento e fornecer descobertas ricas em semântica.

Diante do cenário apresentado, a habilidade de fazer consultas para busca de dados semânticos em terminais SPARQL se destacou amplamente, onde os modelos SKOS e extensão SKOS-XL são primordiais para condicionar tais consultas, juntamente com outras ferramentas e padrões específicos.

Os vocabulários LOD KOS permitem aos produtores de vocabulário criar, manter e enriquecer recursos semânticos reutilizáveis, seguindo os princípios LOD. Esses recursos aproveitam as tecnologias semânticas e ferramentas abertas disponíveis, permitindo a geração rápida de microtesouros e vocabulários satélites por meio de consultas SPARQL. A integração com outros dados e a visualização são

---

<sup>38</sup> Revisitando o Esquema de Dados Abertos de 5 estrelas de Tim Berners-Lee para dados LOD, Zeng e Mayr (2018) descrevem cada classificação, da menor para a maior quantidade de estrelas:

\* Disponível na web (em qualquer formato), mas com uma licença aberta, para ser Dados Abertos.

\*\* Disponível como dados estruturados legíveis por máquina (por exemplo, Excel em vez de uma imagem digitalizada de uma tabela).

\*\*\* Além do item (2), também em um formato não proprietário (por exemplo, CSV em vez de Excel).

\*\*\*\* Todos os itens acima, além disso: Utilize padrões abertos do W3C (RDF e SPARQL) para identificar coisas, para que as pessoas possam referenciar suas informações.

\*\*\*\*\* Todos os itens acima, além disso: Vincule seus dados aos dados de outras pessoas para fornecer contexto (Zeng; Mayr, 2018, tradução livre).

facilitadas por meio de diversos formatos de download. A reutilização desses vocabulários em bancos de dados LOD ou não LOD é uma conquista importante, promovendo a cooperação, reduzindo a duplicação e beneficiando a pesquisa e as cadeias de valor (Zeng; Mayr, 2018).

Os LOD KOS oferecem um potencial infinito para pesquisadores, permitindo a utilização inovadora de vocabulários semanticamente ricos e controlados. A implementação correta e o acesso aos padrões são cruciais para estender a funcionalidade dos SOC além de serem apenas vocabulários controlados. No entanto, ainda há um caminho a percorrer para que os SOC sejam amplamente reconhecidos como bases de conhecimento e ferramentas semânticas. É importante superar as limitações das abordagens tradicionais de pesquisa e navegação na Web, aproveitando ao máximo os dados processáveis por máquina, que são mais poderosos e úteis (Zeng; Mayr, 2018).

Zeng e Mayr (2018) apontam que é fundamental promover uma maior interação entre as comunidades envolvidas na organização do conhecimento e as comunidades envolvidas no desenvolvimento de tecnologias semânticas. Nesse sentido, a participação ativa de pesquisadores que também sejam usuários finais é de valor inestimável. A experiência destes no domínio em que atuam, bem como as necessidades de informação e formas de busca é que estabelecem as perguntas que as bases de conhecimento SOC podem responder. Essa colaboração contribuirá para o desenvolvimento das comunidades de usuários de SOC, ampliando os objetivos e perspectivas de utilização dos SOC de forma diversificada.

Enquanto LOD KOS, os tesouros estão sendo cada vez mais representados em ambientes semânticos da Web. Martínez-González e Alvite-Díez (2019b) afirmam que há iniciativas e grupos de pesquisa trabalhando no desenvolvimento de vocabulários e terminologias LOD KOS como, por exemplo: no campo jurídico, há discussões para construir um modelo de "democracia vinculada"; o SOC Lynx<sup>39</sup>, um sistema que utiliza um grafo de conhecimento jurídico, com terminologias e tesouros, como o Tesouro da Unesco, de modo a oferecer serviços avançados, como registros de documentos legais, definições de termos, classificação de textos e identificação de assuntos; o projeto Terminoteca RDF, que é um protótipo de repositório de

---

<sup>39</sup> *Lynx: Building the Legal Knowledge Graph for Smart Compliance Services in Multilingual Europe.* <http://www.lynx-project.eu/>.

terminologias multilíngues das línguas oficiais da Espanha, que tem como objetivo integrar diversas terminologias em um único grafo unificado, facilitando a pesquisa e a navegação.

Ainda no contexto de LOD KOS, Saliba *et al.* (2021) apresentam o tesauro Zoonom, utilizado para integrar novos termos do domínio de Zoologia a termos existentes no Código Internacional de Nomenclatura Zoológica. Esse tesauro, baseado em SKOS, permitiu que nomes científicos de animais e elementos relacionados fossem controlados em prol da redução da ambiguidade e substituição de termos problemáticos existentes.

O Zoonom foi desenvolvido utilizando o *software* de criação de tesouros de código aberto, Opentheso. O Opentheso está em conformidade com as duas partes da norma ISO 25964. O tesauro é compartilhado online através da plataforma *Linked Open TERminology REsources* (LOTERRÉ). O Zoonom contém 920 termos, excluindo termos das mesmas famílias de palavras, como plurais. Esses termos estão distribuídos em 794 conceitos, 404 etimologias e 58 referências. O Zoonom é dividido em 20 coleções e abrange todos os aspectos da nomenclatura zoológica, desde a nomenclatura teórica até as publicações taxonômicas. É possível encontrar o link para fazer o *download* do arquivo na descrição do Zoonom<sup>40</sup>. Além de ser usado como um glossário clássico, onde é possível pesquisar usando a barra de pesquisa ou navegar em ordem alfabética. O tesauro também possui muitas outras funcionalidades relacionadas à pesquisa e ao controle terminológico (Saliba *et al.*; 2021).

Conhecer e avaliar as funcionalidades de tesouros como o Zoonom é proveitoso, ao passo que profissionais e cientistas da informação tendem a se deparar com demandas por projetos de pesquisa científica, trabalhos corporativos e necessidades organizacionais do funcionalismo público de diferentes níveis de complexidade. Estes níveis influenciam no desenvolvimento e oferta de produtos e serviços de informação estruturados de acordo com parâmetros levantados através de *benchmarking* de boas práticas de organização do conhecimento em incontáveis temáticas.

---

<sup>40</sup> Durante a realização desta pesquisa foram feitas várias tentativas de acesso ao Tesauro Zoonom, porém diferentes erros de acesso foram apresentados.

Percebe-se que, quando se trata do contexto geral de utilização, a identificação das tipologias SOC não são enfocadas, pois os vocabulários controlados, independentemente de serem tesouros ou outro tipo, operam em conjunto para atender a necessidades de uso cujos usuários finais não requerem análise e estudo tipológico de SOC, nem mesmo discriminação específica entre eles. Porém, para o *design*, modelagem, produção, edição e gerenciamento de SOC, a identificação de cada uma de suas tipologias é fundamental, em que esses aspectos discriminatórios embasam as escolhas de ferramentas, padrões, normas, modelos, extensões, bem como as propostas de soluções *ad hoc* capazes de estruturar adequadamente tesouros ou outros SOC interoperáveis e suficientemente eficazes e eficientes para o contexto de uso a que se prestam.

Atualmente, há uma grande diversidade e quantidade de SOC disponíveis para busca e acesso. Isso inclui uma variedade de vocabulários e estruturas de dados relacionadas às diferentes áreas de conhecimento (Rio-Branco, Moreira; Fujita, 2021).

A partir de revisão sistemática, Rio-Branco, Moreira e Fujita (2021) destacam a importância da interoperabilidade e do mapeamento semântico na Web Semântica. O mapeamento é necessário devido à diversidade de sistemas de organização do conhecimento em diferentes áreas. A função principal do mapeamento é alcançar a interoperabilidade entre esses sistemas. A revisão de literatura mostra avanços teóricos, metodológicos, normativos e tecnológicos encorajadores, mas os autores defendem que são necessários mais apoio e conscientização, pois questões técnicas estão interligadas com aspectos sociais e culturais, o que requer políticas informacionais sólidas.

De um modo geral, as fontes de pesquisa utilizadas por Zeng e Mayr (2018) para analisar o cenário de produção e uso de LOD KOS (apresentações, redes sociais e outros registros não formais) indicam que a dinâmica de construção de SOC interoperáveis e conceituais no contexto internacional, em especial os tesouros, é bastante rápida e intensa. Acredita-se que isso se deve aos avanços nas metodologias de produção de SOC semanticamente interoperáveis, bem como à agilidade de aprendizagem e produção de conhecimento propiciada e demandada no contexto Web contemporâneo.

## 6 CONSIDERAÇÕES FINAIS

Esta investigação ocorreu sob o tema representação temática da informação em ambientes digitais, com justificativas que fundamentaram a definição dos objetivos geral e específicos delineados.

Para o presente trabalho de pesquisa, os encadeamentos entre a Ciência da Informação e os SOC do tipo tesouros permitiram fundamentar os conceitos utilizados para análise dos resultados e caracterização do principal objeto de pesquisa investigado: tesouros representados em SKOS.

O enfoque metodológico enquadrado no Modelo Metodológico Quadripolar contextualizou a descrição dos procedimentos técnicos cunhados em benefício da análise de conteúdo categorial de vinte e cinco textos, incluindo a norma ISO 25964 e as recomendações do modelo SKOS. Assim, a análise de conteúdo se firmou em 11 categorias de discussão: 1) relação dos tesouros com outros SOC; 2) conceitos e termos; 3) domínios, microtesouro e grupos de conceitos; 4) relações de equivalência; 5) equivalência composta; 6) relações hierárquicas; 7) Principais termos (*topTerms*) e principais conceitos (*topConcepts*); 8) relações associativas; 9) notas; 10) arranjos; e 11) abordagens de aprimoramento, extensões SKOS e usos do modelo de dados SKOS e da norma ISO 25964.

As justificativas e hipóteses foram embasadas em dois fatos. Sendo que o primeiro – que consiste na utilização ampla do modelo de dados SKOS antes mesmo da publicação da norma ISO 25964 –, de acordo com o *corpus* analisado, intensificou o desenvolvimento de extensões para o modelo. Por seu turno, o segundo – a constatação de que o modelo de dados SKOS contempla todos os tipos de SOC, enquanto a norma ISO 25964 enfoca orientações sobre a construção e mapeamento de tesouros para recuperação da informação –, conforme constatado por meio do estudo do *corpus*, ocasiona similaridades entre as recomendações da norma e as do modelo. A análise dos elementos que compõem essas similaridades poderá embasar e contribuir com pesquisas futuras e com o desenvolvimento e aplicação de tesouros interoperáveis semanticamente em diferentes contextos temáticos na Web. Deste modo, os resultados apresentados têm importância para a vida em uma sociedade cada vez mais ativa na Web, onde o funcionamento de

tesauros em ambientes semânticos favorece a representação informacional em prol da recuperação inequívoca de conteúdos que favoreçam o desenvolvimento social, científico e profissional de diferentes domínios do conhecimento.

Especificamente sobre as justificativas apresentadas de acordo com os contextos de atuação e percepção da autora que antecedem esta pesquisa, constatou-se: breve menção a utilização de tesauros semanticamente interoperáveis no contexto de PLN; que os resultados não apresentaram nenhum indício sobre uso de tesauros para o desenvolvimento de *software* em apoio a desenvolvedores e PO; porém, a relação entre uso de tesauros e enriquecimento semântico no contexto de RI atrelado a páginas Web de diferentes tipos de unidades de informação e escopos temáticos foi detectada no *corpus*; além de unidades de informação, os resultados também apresentam breve indício da representação temática da informação via tesauros aplicados em âmbito empresarial; que ainda que tesauros representados em SKOS e em conformidade com a norma ISO 25964 favoreçam o enriquecimento semântico no contexto de RI, a necessidade de utilização de terminais SPARQL por humanos para consulta aos mesmos torna a interface de consulta não muito amigável e pouco intuitiva. Nesse sentido, a representação de conhecimento registrado contra-hegemônico via tesauros semanticamente interoperáveis se demonstra mais complexa no tangente a sua utilização por comunidades indígenas e afro-brasileiras, por exemplo, para as quais os recursos para desenvolvimento, sensibilização e treinamento ao uso de SOC tende a ser menos acessíveis do que para a comunidade jurídica hegemonicamente estabelecida, por exemplo.

Na ocasião da formulação do objetivo geral desta pesquisa, acreditava-se que o modelo SKOS era uma alternativa de baixo custo para construção de SOC, por ser de acesso livre. Porém, após a análise dos resultados, considera-se que aspectos como o tempo, conhecimento teórico e técnico exigidos para a modelagem de um tesouro em SKOS, em conformidade com a norma ISO 25964, precisam ser analisados para embasar a formulação de estratégias que visem o alcance de investimentos adequados para *design*, modelagem, construção, gerenciamento e edição de SOC, em especial os tesauros enquanto LOD KOS. Visto que soluções de representação temática, ágeis, eficientes, eficazes e intuitivas exigem trabalho de estruturação e manutenção minucioso.

Acredita-se que tanto o objetivo geral como os objetivos específicos foram alcançados, visto que os resultados aqui apresentados estão em consonância com a delimitação proposta, qual seja: analisar a estrutura do modelo de dados SKOS, recomendações e norma relacionada, atentando-se para especificidades e similaridades com a estrutura de tesouros clássicos; apontar aspectos normativos em conformidade com a norma ISO 25964 para a representação de tesouros conceituais e interoperáveis; e verificar abordagens de aprimoramento, extensões e usos no âmbito do cenário geral de aplicações conjuntas do modelo SKOS e da norma ISO 25964.

Um aspecto que se destacou amplamente no âmbito dos resultados foi o fato de tanto o modelo de dados SKOS quanto a norma ISO 25964 serem compatíveis entre si, mesmo com respectivos lançamentos e inícios de aplicação em momentos diferentes. Esse aspecto torna a norma aplicável mediante a flexibilidade do SKOS em prol de suprir diversas necessidades de elaboração de tesouros e outros SOC em ambiente Web, em benefício de contextos de uso específicos. E esses usos específicos se refletem não apenas na abordagem temática da informação como também na não neutralidade dos tesouros, que tendem a refletir os interesses que permeiam o domínio temático que se prestam a representar. Os usos específicos também levam às necessidades precisas de construção dos tesouros aqui abordados, onde as soluções *ad hoc* estão sempre presentes. Deste modo, surgem diferentes extensões, *softwares* e outras tecnologias aptas a sustentar o modelo SKOS e/ou a norma ISO 25964, bem como outros padrões e instrumentos relacionados à construção de ambientes semânticos na Web. Neste cenário, a colaboração ocorre por meio da interoperabilidade e mapeamento a nível global.

Os resultados alcançados evidenciam que a maioria das publicações do *corpus* de análise se debruçaram sobre a estrutura da norma ISO 25964 para relacioná-la ao SKOS, atentando-se para aspectos conceituais e notacionais e, na maioria dos casos, demonstrando exemplos.

Conforme apresentado na subseção 5.12.4, mesmo que a Ciência da Informação aborde modelos, conceitos e teorias, é importante que seus pesquisadores e profissionais compreendam as relações entre as partes técnicas e tecnológicas voltadas à aplicação dos conceitos e teorias atrelados à modelagem

contemporânea de SOC como os tesouros LOD KOS. Tal compreensão relacional favorece a comunicação científica e mercadológica entre profissionais e pesquisadores da Ciência da Informação e da Ciência da Computação, por exemplo, beneficiando a oferta de produtos e serviços de informação baseados em SOC do tipo tesouros conceituais e interoperáveis em ambiente Web. É a partir deste entendimento que os textos da área da computação foram considerados e mantidos para análise, pois apresentaram discussões sobre conceitos identificados que vão ao encontro dos textos oriundos da Ciência da Informação.

Neste trabalho foram abordadas predominantemente as semelhanças entre o modelo de dados SKOS e a norma ISO 25964. Esta que por sua vez é tratada como modelo ou até mesmo ontologia por alguns autores. Neste sentido, um estudo mais amplo, mas que inclua a análise das diferenças entre a norma e o modelo a partir da construção de outros tipos de SOC, se faz necessário. A começar pelo aprofundamento experimental das comparações de estrutura e condições de integridade na norma ISO 25964 e do modelo SKOS, que expressa restrições de integridade por meio de axiomas.

### **6.1 Limitações da pesquisa**

O estudo da norma ISO 25964 foi fundamental para o desenvolvimento desta pesquisa. Porém, o acesso à mesma foi delicado, pois o Sistema Integrado de Bibliotecas da UFSCar não dispunha do acesso completo. Assim, foi necessário buscar acesso em outras bibliotecas, até que a Biblioteca da ECA/USP acolheu tal demanda informacional, disponibilizando o acesso da norma para estudo.

A escassez de publicações nacionais e internacionais para análise foi um limitador. Como a pesquisa contemplou bases de dados nacionais e internacionais, causou estranhamento o relativo silêncio nacional, na literatura científica, sobre as similaridades entre o modelo SKOS e a norma ISO 25964, bem como as aplicações dessas similaridades. Acredita-se que o número de publicações sobre as relações entre o modelo SKOS e a norma ISO 25964 não seja maior pelo fato de os contextos de análise e uso, da norma e do modelo, serem bastante dinâmicos, conforme discutido na subseção 5.12.4.

Relativo à interpretação de alguns dos termos específicos em outros idiomas, ressalta-se que as publicações nacionais analisadas (Barbosa; Viera, 2022; Berti Junior, 2017; Rio-Branco; Moreira; Fujita; 2021;) foram fundamentais para embasar a compreensão de todo o *corpus*, pois alguns exemplos de termos e conceitos apresentados em inglês e espanhol muitas vezes não tinham correspondentes exatos em português e, se tinham, a sua tradução não era tão simples.

Outra limitação deste trabalho é que não foi verificada a continuidade de pesquisa, por parte dos autores internacionais, sobre a norma ISO 25964 e o modelo de dados SKOS para representação de tesouros ou outros SOC, conforme os resultados obtidos a partir da pesquisa bibliográfica.

## **6.2 Sugestões de pesquisas futuras e complementares**

Considera-se que ainda há muito o que investigar dentre os aspectos que permeiam as similaridades entre a norma ISO 25964 e o modelo de dados SKOS, pois esta pesquisa está longe de ter esgotado todos os aspectos de similaridades e divergências entre a norma e o modelo. Deste modo, este trabalho apresenta as seguintes sugestões de pesquisas futuras e complementares:

- Aplicação de um procedimento metodológico que viabilize, primeiramente, a busca por documentos que tratam da norma ISO 25964 para depois selecionar os documentos que abordam o SKOS. Esta manobra pode resultar na identificação de publicações mais recentes, visto que a norma surge depois da recomendação formal do modelo de dados SKOS;
- Investigação das relações de mapeamento entre tesouros e os demais tipos de SOC a partir das instruções da norma ISO 25964;
- Investigação sobre a possibilidade de representar tesouros apenas a partir do modelo de dados ISO 25964, sem a modelagem em SKOS, visto que ambos são considerados modelos de dados por alguns autores;
- Estudo detalhado e comparativo de componentes semânticos de diversos tipos de SOC, incluindo os tesouros;
- Revisão bibliográfica sobre as diferenças e similaridades entre ontologias e tesouros representados em SKOS e de acordo com a norma ISO 25964;

- Realização de experimentos, de utilização de tesouros e outros SOC representados em SKOS de acordo com a norma ISO 25964, em *softwares* como Tematres, VocBench, SKOSEd e PoolParty;
- Estudo sobre a existência real da composicionalidade das relações de transitividade das relações hierárquicas, equivalentes e associativas de tesouros representados em conformidade com a norma ISO 25964;
- Estudo sobre a etimologia das palavras *Thesaurus* e *Thesauri*, por vezes traduzidos como dicionário de sinônimos ou tesouro (a partir do latim *thesauru*) também poderia revitalizar a conceituação do termo tesouro e, assim, contribuir para a base teórico-conceitual da Ciência da Informação;
- Estudo sobre condições de integridade entre a norma ISO 25964 e o modelo SKOS;
- Estudo que trate sobre o que a norma ISO 25964 trouxe de novo em relação às normas anteriores, de modo a analisar como os principais tesouros brasileiros eram representados antes da norma e atualmente, com a existência da norma;
- Estudo para avaliar a situação atual das abordagens de aprimoramento, extensões SKOS e usos apresentados na subseção 5.12; e
- Análise teórica sobre os tipos de modelos de dados existentes e relacionados à Organização e Representação do Conhecimento no contexto da Ciência da Informação.

Por fim, considera-se que o estudo realizado confere aportes teóricos e metodológicos para *design*, construção, migração, atualização, manutenção e gerenciamentos de tesouros conceituais e interoperáveis semanticamente em ambiente Web.

## REFERÊNCIAS

ALEXIEV, Vladimir; ISAAC, Antoine; LINDENTHAL, Jutta. On the composition of ISO 25964 hierarchical relations (BTG,BTP, BTI). **International Journal on Digital Libraries**, v. 17, p. 39-48, 2016

ALMEIDA, Daniela Pereira dos Reis de; ANTONIO, Deise Maria; BOCCATO, Vera Regina Casari; GONÇALVES, Maria Carolina; RAMALHO, Rogério Aparecido Sá; Paradigmas contemporâneos da Ciência da Informação: a recuperação da informação como ponto focal. **Revista Eletrônica Informação e Cognição**, v.6, n.1, p.16-27, 2007. Disponível em: <https://brapci.inf.br/index.php/res/download/144585>. Acesso em: 10 mar. 2021

ALVITE-DÍEZ, María -Luisa; MARTÍNEZ-GONZÁLEZ, Mercedes. Representación de tesauros en SKOS: tendencias y desafíos para su integración en la web wemántica. **Scire**, v. 27, n. 2, jul./dez., p. 13-20 2021

ARAÚJO, Carlos Alberto Ávila. Teorias e tendências contemporâneas da Ciência da Informação. **Informação em Pauta**, Fortaleza, v. 2, n. 2, p. 9-34, dez. 2017. Disponível em: <http://www.periodicos.ufc.br/informacaoempauta/article/view/20162/71589>. Acesso em 16 dez. 2020

ASTM COMPASS. **ISO 25964-1:2011**. Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval. 2022

ASTM COMPASS. **ISO 25964-2:2013**. Information and documentation — Thesauri and interoperability with other vocabularies — Part 2: Interoperability with other vocabularies. 2022

BAEZA-YATES, Ricardo; RIBEIRO NETO, Berthier. **Recuperação de informação: conceitos e tecnologias das máquinas de busca**. 2. ed. Porto Alegre: Bookman, 2013.

BALKAN, Lorna; BELL, Lucy. Linking Thesauri - ELSST as a Hub for Social Science Data Terms. **IASSIST Quarterly**, v. 38, n. 2, 2014

BANDYOPADHYAY, Subarna; MUKHOPADHYAY, Parthasarathi. Developing SKOS-compliant multilingual thesaurus: an ISO 25964 based Approach. **Information studies**, v. 21, n. 4, 2015

BARBOSA, Alice Príncipe. **Teoria e prática dos sistemas de classificação bibliográfica**. Rio de Janeiro: IBBD, 1969. 441 p.

BARBOSA, Everton Rodrigues; VIERA, Angel Freddy Godoy. Relações semânticas e interoperabilidade em tesauros representados em skos. **Informação & Informação**, v. 27, n. 1, p. 229-252, 2022. DOI: 10.5433/1981-8920.2022v27n1p229. Acesso em: 19 jan. 2023.

BARRETO, Aldo de Albuquerque. A condição da informação. **São Paulo em Perspectiva**, v. 16, n. 3, p. 67-74, 2002

BARROS, Lidia Almeida. **Curso básico de terminologia**. São Paulo: Edusp, 2004. 296 p.

BERTI JUNIOR, Decio Wey; SOERGEL, Dagobert; LIMA, Gercina Ângela de; MACULAN, Benildes Coura dos Santos. Semiautomatização de relações em tesouros: uma proposta para o refinamento de relacionamentos semânticos a partir do tesouro agrovoc. **Informação & Informação**, v. 22, n. 3, p. 377-404, 2017. DOI: 10.5433/1981-8920.2017v22n3p377. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/28724/22738>. Acesso em: 19 jan. 2023.

BEZARES, Gonzalo Mochón; RODRÍGUEZ, Eva María Méndez; FUENTE, Gema Bueno de la. Conversión normalizada (SKOS) de sistemas de organización del conocimiento interoperables en la web. **Investigación Bibliotecológica**, v. 34, n. 82, jan./mar., 2020

BRUYNE, Paul de; HERMAN, Jacques; SCHOUTHEETE, Marc de. **Dinâmica da pesquisa em ciências sociais: os pólos da prática metodológica**. Rio de Janeiro: Livraria Francisco Alves Editora S.A., 1977.

BURKE, P. Dentzien, Plinio (trad). **Uma historia social do conhecimento: de Gutenberg a Diderot**. Rio de Janeiro: Jorge Zahar Editor, 2003.

BUSH, Vannevar. **As we may think**. Atlantic Monthly, v. 176, n.1, p.101-108, 1945

CABRÉ, Maria Tereza. La terminología hoy: concepciones, tendencias y aplicaciones. **Ciência da Informação**, v. 24, n. 3, 1995. Disponível em: <https://revista.ibict.br/ciinf/article/view/567/568>. Acesso em: 9 mar. 2022

CABRÉ, Maria Tereza. **La terminología: teoría, metodología, aplicaciones**. Barcelona: Ed.Antártida; Empúries, 1993.

CABRÉ, Maria Tereza. Terminología Y documentación. In: \_\_\_\_\_ **La terminología: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos**. Barcelona: ULA/Universitat Pompeu Fabra, 1999. p. 231-247

CAMPOS, Maria Luiza de Almeida. **Linguagem documentária: teorias que fundamentam sua elaboração**. Niterói: EdUFF, 2001. 133p.

CAPURRO, Rafael. Epistemologia e Ciência da Informação. In: **Encontro Nacional de Pesquisa em Ciência da Informação**, 2003, Belo Horizonte, Anais... Belo Horizonte: ENANCIB, 2003, p. 1-21. Disponível em: [http://www.capurro.de/enancib\\_p.htm](http://www.capurro.de/enancib_p.htm). Acesso em: 13 mar 2020

CARACCIOLO, Caterina; STELLATO, Armando; MORSHED, Ahsan; JOHANNSEN, Gudrun; RAJBAHNDARI, Sachit; JAQUES, Yves; KEIZER, Johannes. The AGROVOC Linked Dataset. **Semantic Web journal: Special Call for Linked Dataset descriptions**, 2012. Disponível : <https://www.semantic-web-journal.net/content/agrovoc-linked-dataset>. Acesso em 17 jul. 2022

CASTRO, Fabiano Ferreira de; JESUS, Ananda Fernanda de. Adequando dados bibliográficos ao linked data: requisitos necessários, vantagens e desafios.

**Informação & Sociedade**, v. 28, n. 3, p. 149-168, set./dez., 2018. Disponível em: <https://brapci.inf.br/index.php/res/download/109618>. Acesso em: 13 ago. 2021

CATARINO, Maria Elisabete; CERVANTES, Brígida Maria Nogueira, ANDRADE Ilza Almeida de. A representação temática no contexto da web semântica. **Informação & Sociedade**, v.25, n.3, p. 105-116, set./dez. 2015

CATARINO, Maria Elisabete. Simple Knowledge Organization System: construindo sistemas de organização do conhecimento no contexto da web semântica. **Informação & Tecnologia (ITEC)**, v. 1, n. 1, p. 17-28, jan./jun., 2014. Disponível em: <https://periodicos.ufpb.br/index.php/itec/article/view/19307/10980>. Acesso em 20 dez. 2020

CERVANTES, Brígida Maria Nogueira; RAMALHO, Rogério Aparecido Sá; GONÇALEZ, Paula Regina Ventura Amorim; SANTOS, José Carlos Francisco dos. Representação e Recuperação da Informação na Web: aspectos teóricos e tecnológicos. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 13, p. 409-426, 2018. Disponível em: <https://brapci.inf.br/index.php/res/download/108521>. Acesso em: 14 jan. 2021

CHAVES, F N. Oralidade, memórias e resistência cultural popular. 2018. **JAMAXI Revista de História e Humanidades**, Rio Branco, v. 2, n. 1, p. 52-61, mar. 2018. Disponível em: <https://periodicos.ufac.br/index.php/jamaxi/article/view/1714>. Acesso em: 25 dez. 2020

CINTRA, Anna Maria Marques; TÁLAMO, Maria de Fátima Gonçalves Moreira; LARA, Marilda Lopes Ginez de; KOBASHI, Nair Yumiko. **Para entender as linguagens documentárias**. 2 ed. rev e ampl. São Paulo, Polis, 2002. 92 p.

CLARKE, Stella G. Dextre; ZENG, Marcia Lei. From ISO 2788 to ISO 25964: The Evolution of Thesaurus Standards towards Interoperability and Data Modeling. In: **Information Standards Quality**, Special Edition: Year in Review and State of the Art of the Standards, p.20-26, 2012.

CONEGLIAN, C. S. **Recuperação da Informação com abordagem semântica utilizando Linguagem Natural**: a Inteligência Artificial na Ciência da Informação. Orientador: José Eduardo Santarém Segundo. 2020, 195 f. - Tese (Tese de Doutorado em Ciência da Informação) – Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), de Filosofia e Ciências, Marília, 2020.

CURRÁS, Emilia. **Ontologías, taxonomía y tesauros**: manual de construcción y uso. 3. ed. Gijón, Trea, 2005. 337 p.

CURRÁS, Emilia; ROBREDO, Jaime (trad). **Ontologias, taxonomia e tesouro em teoria de sistemas e sistemática**. Brasília: Thesaurus, 2010. 182 p.

CURRÁS, Emilia. **Tesauros**: linguagens terminológicas. Brasília: IBICT, 1995.

DAHLBERG, Ingetraut. A referent-oriented, analytical concept theory of Interconcept. **International Classification**, v. 5, n. 3, p. 122-151, 1978. Disponível em: <https://www.nomos-elibrary.de/10.5771/0943-7444-1978-3-142.pdf>. Acesso em 20 ju.I 2022

DODEBEI, Vera Lúcia Doyle. **Tesouro**: linguagem de representação da memória documentária. Rio de Janeiro: Intertexto; Interciência, 2002

FAULSTICH, Enilde. A socioterminologia na comunicação científica e técnica.

**Ciência e Cultura**, v. 58, n. 2, p. 27-31, abr./jun. 2006. Disponível

em:[http://cienciaecultura.bvs.br/scielo.php?script=sci\\_arttext&pid=S0009-6725200600200012](http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S0009-6725200600200012). Acesso em: 10 jan. 2022

FERNEDA, E.; DIAS, G. A. OntoSmart: um modelo de recuperação de informação baseado em ontologia. **Perspectivas em Ciência da Informação**, v. 22, n. 2, p.

170-187, jun. 2017. Disponível em: <http://portaldeperiodicos.eci.ufmg.index.php?pciarticleview/2081/1882>. Acesso em: 16 ago. 2022

FERRAZ, Reinaldo. **Tendências da web**. [S.l.]: Senac, 2018.

FERREIRA, Jaider Andrade; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. O modelo de dados Resource Description Framework (RDF) e o seu papel na

descrição de recursos. **Informação & Sociedade: Estudos**, v. 23, n. 2, p. 13-23, 2013. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/15436>.

Acesso em: 21 jan. 2020.

FOSKETT, Antony Charles; LEMOS, Antonio Agenor Briquet de (trad.). **A abordagem temática da informação**. São Paulo: Polígono; Brasília: Editora Universidade de Brasília, 1973. 437p.

FUJITA, Mariângela Spotti Lopes; MOREIRA, Walter (orgs.). **Manual do planejamento, construção e manutenção do tesauro Unesp para bibliotecas**: do conceitual ao prático. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2021

GARDIN, Jean-Claude. Document analysis and linguistic theory. **Journal of Documentation**, v. 29, n.2, p.137-168, 1973

GARDIN, Jean-Claude. Eléments d'un modèle pour la description des lexiques documentaires. **Bulletin des Bibliothèques de France**, n. 5, p.171-182, 1966

GARDIN, Jean-Claude. Procédures d'analyse sémantique dans les sciences humaines. In: POUILLON, J., MARANDA, P. **Échanges et communications**: mélanges offerts à Claude Lévi-Strauss à l'occasion de son 60ème anniversaire. Paris, Mouton, p.628-657, 1970

GEBAUER, Gunter. **O pensamento antropológico de Wittgenstein**. São Paulo: Loyola, 2013. 223p.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (Org.). **Métodos de pesquisa**. Porto Alegre: Ufrgs, 2009. Disponível em:

<http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf>. Acesso em: 07 jan. 2021

GRACIOSO, Luciana de Souza; SALDANHA, Gustavo Silva. Investigando a significação em Wittgenstein. In: GRACIOSO, L. S.; SALDANHA, G S. **Ciência da Informação e Filosofia da Linguagem: da pragmática informacional à web pragmática**. Araraquara: Junqueira & Marin, 2011. Disponível em:

[https://books.google.com.br/books?id=B7Z2DwAAQBAJ&pg=PA11&hl=pt-BR&source=gbs\\_selected\\_pages&cad=2#v=onepage&q&f=false](https://books.google.com.br/books?id=B7Z2DwAAQBAJ&pg=PA11&hl=pt-BR&source=gbs_selected_pages&cad=2#v=onepage&q&f=false). Acesso em: 07 jul. 2021

GRACIOSO, Luciana de Souza. **Sobre um outro lugar da linguagem nos estudos da informação**. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Salvador, BA. Anais... Salvador: UFBA, 2016. Disponível em:

<http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/3073/2016-GT01-CO-07.pdf?sequence=1>. Acesso em: 14 jul. 2021

GUIMARÃES, José Augusto Chaves. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). **Revista Ibero-Americana de Ciência da Informação**, Brasília, v. 1, n. 1, p. 77-99, jan./jun. 2008. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/940/815>. Acesso em: 28 mar. 2023

GUIMARÃES, José Augusto Chaves. Abordagens teóricas de tratamento temático da informação (TTI): catalogação de assunto, indexação e análise documental. **Ibersid**, Zaragoza, p. 105-117, 2009. Disponível em: <https://brapci.inf.br/index.php/res/download/196885>. Acesso em: 28 mar. 2023

HJØRLAND, Birger; ALBRECHTSEN, Hanne. Toward a new horizon in information science: domain-analysis. **Journal of the American Society for Information Science**, v. 46, n. 6, p. 400-425, 1995.

HJØRLAND, Birger. Knowledge organization. In: **ISKO Encyclopedia of Knowledge Organization**, Knowledge organization, v. 43, n. 6, p. 475-484, 2016. Disponível em: [https://www.isko.org/cyclo/knowledge\\_organization](https://www.isko.org/cyclo/knowledge_organization). Acesso em: 11 mar. 2022

HJØRLAND, Birger. "Subject (of documents)". **Knowledge Organization**, Würzburg, v. 44, n. 1, p. 55-64, 2017. Disponível em: <https://www.isko.org/cyclo/subject>. Acesso em: 21 jul. 2022.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 25964**: Information and documentation: thesauri and interoperability with other vocabularies - part 2: interoperability with other vocabularies. Genebra, 2013.

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

ISAAC, Antoine; SUMMERS, E. (ed.). **SKOS Simple Knowledge Organization System Primer**. W3C Working Group Note, 18 August 2009. Disponível em: <https://www.w3.org/TR/skos-primer/>. Acesso em: 19 set 2019.

ISOTANI, Seiji; BITTENCOURT, Ig Ibert. **Dados Abertos Conectados**: em busca da web do Conhecimento. [s.l.]: Novatec Editora, 2015. Disponível em: [http://pgcl.uenf.br/arquivos/dadosabertosconectados\\_011120181613.pdf](http://pgcl.uenf.br/arquivos/dadosabertosconectados_011120181613.pdf). Acesso em: 22 jun. 2022

ISO TC46/SC9/WG8; ISAAC, Antoine. **Correspondence between ISO 25964 and SKOS/SKOS-XL models**. 2012

JESUS, Samantha Augusta dos Santos de; MOREIRA, Walter. Contribuições teóricas da terminologia nos sistemas e processos de organização do conhecimento.

**Múltiplos Olhares em Ciência da Informação**, v. 8 No. 2, n. 2, 2018. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/136759>. Acesso em: 11 mar. 2022.

KOBASHI, Nair Yumiko. Organização da informação e do conhecimento: aspectos informacionais e comunicacionais. In: **ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO**, 17., 2016, Salvador. Anais... Salvador: UFBA, 2016. Disponível em:

[http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/3387/2016\\_GT2-CO\\_14.pdf?sequence=1](http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/3387/2016_GT2-CO_14.pdf?sequence=1). Acesso em: 14 jul. 2021

LANCASTER, Frederick Wilfrid. **Indexação e resumos: teoria e prática**. 2. ed. rev. atual. Brasília, DF: Briquet de Lemos/Livros, 2004.

LARA, Marilda Lopes Ginez de. Conceitos de organização e representação do conhecimento na ótica das reflexões do grupo tema. **Informação & Informação**, v. 16; n. 3; p. 92-121, jan./jun. 2011. Disponível em:

<https://brapci.inf.br/index.php/res/download/45240>. Acesso em: 12 mar. 2022

LARA, Marilda Lopes Ginez de. Documentary languages and knowledge organization systems in the context of the semantic web. **TransInformação**, v. 25, n. 2, p. 145-150, maio/ago, 2013. Disponível em:

<https://www.scielo.br/j/tinf/a/GRvkDGNqc6NYQtBx57D6cDy/?format=pdf&lang=en>. Acesso em: 20 maio 2022

LAUFER, Carlos. **Guia de web semântica**. Governo do Estado de São Paulo e Governo do Reino Unido, 2015. Disponível em:

<http://www.governoaberto.sp.gov.br/wp-content/uploads/2016/05/Book-Web-Semantica.pdf>. Acesso em: 19 mar. 2022

LE COADIC, Yves-François; GOMES, Maria Yêda F. S. de Filgueiras (trad.). **A ciência da informação**. 2. ed. Brasília, Briquet de Lemos, 2004. 124 p.

LIMA, Juliana Soares; OLIVEIRA, Thiago Pinheiro Ramos de; WANDERLEY, Ana Isabel Ferreira. **Método Quadripolar**. Fortaleza: Programa de Pós-graduação em Ciência da Informação, 2018. 22 slides. Disponível em: <http://bit.ly/2IsmiVZ>. Acesso em: 07 dez. 2022.

LISPECTOR, Clarice. **A paixão segundo G. H.** Rio de Janeiro: Rocco, 2014 [1964]. 191p.

MACHADO, Diego de Queiroz; MATOS, Fátima Regina Ney; SENA, Augusto Marcos Carvalho de; BAPTISTA, Maria Manuel. O Modelo Metodológico Quadripolar de Bruyne, Herman e Schoutheete e as pesquisas qualitativas de fenômenos sociais. **Investigação Qualitativa em Ciências Sociais**, v. 3, p. 276-285, 2016. Disponível

em: <https://proceedings.ciaiq.org/index.php/ciaiq2016/article/view/952/935>. Acesso em: 03 dez. 2022.

MACHADO, Raildo de Souza; ZAFALON, Zaira Regina. **Catálogo**: dos princípios e teorias ao RDA e IFLA. João Pessoa: Editora UFPB, 2020. 128 p. ISBN 978-65-5942-000-1

MACULAN, Benildes Coura Moreira dos Santos; LIMA, Gercina Angela Borém de Oliveira. Buscando uma definição para o conceito de “conceito”. **Perspectivas em Ciência da Informação**, v. 22, n. 02, abr./jun. 2017

MANNING, Christopher D. ; RHAGAVAN, Prabhakar; SCHUTZE, Hinrich. **Introduction to information retrieval**. Cambridge: Cambridge University Press, 2008.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Fundamentos de Metodologia Científica**. São Paulo: Atlas, 2022.

MARTÍNEZ-GONZÁLEZ, Mercedes M; ALVITE-DÍEZ, María-Luisa. A semantic web methodological framework to evaluate the support of integrity in thesaurus tools. **Journal of Information Science**, p. 1-14, 2019a

MARTÍNEZ-GONZÁLEZ, Mercedes M; ALVITE-DÍEZ, María-Luisa. Thesauri and Semantic Web: Discussion of the evolution of thesauri toward their integration with the Semantic Web. **IEEE Access**, p. 153151-153170, v.20, 2019b

MARTÍNEZ-GONZÁLEZ, Mercedes M; ALVITE-DÍEZ, María-Luisa. The support of constructs in thesaurus tools from a Semantic Web perspective: Framework to assess standard conformance. **Computer Standards & Interfaces**, v. 65, p. 79-91, 2019c

MATTHEWS, B.; MILES, A.; WILSON, M. **Modelling thesauri for the Semantic Web**: Semantic Web Advanced Development for Europe (SWAD-Europe), 2009. Disponível em: <http://www.webcitation.org/5m2lmCyQY>. Acesso em: 15 maio 2022

MILES, Alistair; BECHHOFFER, Sean. **SKOS Simple Knowledge Organization System Reference**. W3C Recommendation, 2009. Disponível em: <http://www.w3.org/TR/skos-reference/>. Acesso em: 20 jun 2022

MORAIS, Marília Winkler de; RAMALHO, Rogério Aparecido Sá. Representação em SKOS de um microtesauro de conhecimentos estratégicos nas organizações. **Perspectivas em Ciência da Informação** (Impresso), v. 24, p. 187-198, 2019.

MOREIRA, Walter. Relações conceituais como ponto de inflexão entre linguagens documentais, terminologia e ontologias. **Scire**, v. 18, n. 2, p. 123-127, jul./dez., 2012

NATIONAL INFORMATION STANDARDS ORGANIZATION. **ISO 25964**: the international standard for thesauri and interoperability with other vocabularies. Baltimore, 2023. Disponível em: <https://www.niso.org/schemas/iso25964>. Acesso em: 17 ago 2022.

NOTH, Winfried. **Handbook of Semiotics**. Bloomington: Indiana University Press, 1995.

NOTH, Winfried. **Panorama da semiótica**: de Platão a Pierce. 4. ed. São Paulo, Annablume, 2005.

OTLET, Paul. **Documentos e Documentação**. Trad. de Hagar Espanha. Paris, 1937. Introdução aos trabalhos do Congresso Mundial da Documentação Universal, realizado em Paris, em 1937. Disponível em: <http://www.conexaorio.com/bit/otlet/index.htm>. Acesso em: 9 abr. 2021.

OTLET, Paul. The science of bibliography and documentation. 1903. In: RAYWARD, W. Boyd. **The international organization and dissemination of knowledge**: Selected essays of Paul Otlet. Amsterdam: Elsevier, 1990

PANDO, Daniel Abraão. **Epistemologia da organização da informação**: uma análise de sua cientificidade no contexto brasileiro. 2018. 463 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2018. Disponível em: [https://repositorio.unesp.br/bitstream/handle/11449/153395/pando\\_da\\_dr\\_mar.pdf?sequence=3&isAllowed=y](https://repositorio.unesp.br/bitstream/handle/11449/153395/pando_da_dr_mar.pdf?sequence=3&isAllowed=y). Acesso em: 27 jul. 2020.

PASTOR-SÁNCHEZ, Juan-Antonio. ISO-THES: ampliando Skos a partir de la norma de tesauros ISO 25964. **Anuario ThinkEPI**, v. 7, p. 189-193, 2013a

PASTOR-SÁNCHEZ, Juan-Antonio. Marcado semántico: tecnologías y aplicación para la representación de sistemas de organización del conocimiento en el contexto linked open data. **Scire**, v. 19, n. 2, jul./dez., p. 55-68, 2013b

PASTOR-SÁNCHEZ, Juan-Antonio; MARTÍNEZ-MÉNDEZ, Francisco-Javier; RODRÍGUEZ-MUÑOZ, José-Vicente. Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data. **El profesional de la información**, v. 21, n. 3, p. 245-253, maio-jun. 2012.

PASTOR-SANCHEZ, Juan-Antonio; MARTÍNEZ-MENDEZ, Francisco Javier; RODRIGUEZ-MUÑOZ, José Vicente. Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. **Information Research**, v. 14, n. 4, p. 1-15, 2009. Disponível em: <http://www.informationr.net/ir/14-4/paper422.html>. Acesso em: 12 jan. 2023.

PASTOR-SÁNCHEZ, Juan-Antonio. Proposal to represent the UNESCO Thesaurus for the semantic web applying ISO-25964. **Brazilian Journal of Information Science**, v. 9, n. 2, p. 1-8, 2016. Disponível em: <https://doi.org/10.36311/1981-1640.2015.v9n2.01.p1>. Acesso em: 22 jan. 2023

PEIRCE, Charles Sanders. **Semiótica**. 2ª ed. São Paulo: Perspectiva, 1995

PETTER, Margarida Maria Taddoni. Linguagem, língua, lingüística. In: José Luiz Fiorin. (Org.). **Introdução à lingüística**. 2ª ed. São Paulo: Contexto, 2010, v.1, p. 11-24.

PINHEIRO, Lena Vania Ribeiro; FERREZ, Helena Dodd. **Tesauro Brasileiro de Ciência da Informação**. Rio de Janeiro; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), 2014. E-book. Disponível em: <http://sitehistorico.ibict.br/publicacoes-e-institucionais/tesauro-brasileiro-de-ciencia-d>

a-informacao-1/copy\_of\_TESAUROCOMPLETOFINALCOMCAPA24102014.pdf.  
Acesso em: 17 mai. 2022.

RAMALHO, Rogério Aparecido de Sá. Análise do Modelo de Dados SKOS: Sistema de Organização do Conhecimento Simples para a Web. **Informação & Tecnologia (ITEC)**, v. 2, n. 1, p. 66-79, jan./jul., 2015. Disponível em: <https://periodicos.ufpb.br/index.php/itec/article/view/25995>. Acesso em 03 jan. 2021

RAMALHO, Rogério Aparecido Sá. Representação SKOS da categoria Tecnologia da Informação e Comunicação do Tesouro Brasileiro de Ciência da Informação. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 10, p. 1-13, 2017. Disponível em: <https://revistas.ancib.org/index.php/tpbci/article/view/417/416>. Acesso em: 15 jan. 2021

RAMALHO, Rogério Aparecido Sá; SOUSA, Janailton Lopes. Diretrizes para avaliação de sistemas de organização do conhecimento representados em SKOS. **Informação & Informação (Online)**, v. 24, p. 126-138, 2019. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/37986/pdf>. Acesso em 10 jan. 2021

RANGANATHAN, Shiyali Ramamrita. **As cinco leis da biblioteconomia**. Brasília: Briquet de Lemos / Livros, 2009. 336 p.

RAUTENBERG, Sandro; SOUZA, Lucélia de; KELNIAR, João Pedro. Uma proposta de tesauro para a ciência da computação: organização de elementos de conhecimento com SKOS. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 26, p. 01-15, 2020.

RAYWARD, W. Boyd. The origins of information science and the international institute of bibliography/international federation for information and documentation (FID). **Journal of the American Society for Information Science**, v. 48, n. 4, p. 289-300. 1997

RIO-BRANCO, Luciana Beatriz Piovezan; MOREIRA, Walter; FUJITA, Mariângela Spotti Lopes. Mapeamento entre sistemas de organização do conhecimento: revisão sistemática de literatura. **Informação em Pauta**, v. 6, p. 1-21, 2021.

ROBREDO, Jaime. **Da ciência da informação revisitada aos sistemas humanos de informação**. Brasília: Thesaurus; SSRR Informações, 2003. 245p.

BARITÉ, Mario. Sistemas de Organización del conocimiento: una tipología actualizada. **Informação & Informação**, v. 16, n. 3, p. 122-139, jan./jun. 2011

SALIBA, Elie Mario; LEBBE, Régine Vignes; Dubois, Alain; OHLER, Annemarie. Zoonom: Gathering the concepts of zoological nomenclature in an electronic thesaurus. **Biodiversity Information Science and Standards**, 2021

SANTOS, José Carlos Francisco; MOREIRA, Walter. SKOS: uma análise sobre as abordagens e suas aplicações na Ciência da Informação. **Informação & Informação**, Londrina, v. 23, n. 3, p. 362-389, set./dez. 2018. Disponível em <http://www.uel.br/revistas/uel/index.php/informacao/article/view/29303>. Acesso em: 20 jan. 2021

SANTOS, Plácida Leopoldina Ventura Amorim da Costa; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Perspectivismo e tecnologias de informação e comunicação: acréscimos à Ciência da Informação. **DataGramaZero: revista de Ciência da Informação**, v. 10, n. 3, p. 1-9, jun. 2009. Disponível em: <https://brapci.inf.br/index.php/res/download/45167>. Acesso em 17 dez. 2020

SARACEVIC, Tefko. Information Science. **Journal of the American Society for Information Science**. v. 50, n. 12, p. 1051–1063, out. 1999

SARACEVIC, Tefko. Information science. In: BATES, Marcia J.; MAACK, Mary Niles (ed.). **Encyclopedia of Library and Information Science**. New York: Taylor & Francis. p. 2570-2586, 2009. Disponível em: <https://tefkos.comminfo.rutgers.edu/SaracevicInformationScienceELIS2009.pdf>. Acesso em 18 dez. 2021

SEVERINO, Antônio Joaquim. **Metodologia do Trabalho Científico**. São Paulo: Cortês Editora, 2009.

SHINTAKU, Milton; SABBAG, Deise Maria Marcos; COSTAL, Marcelle; MENÊSES, Raíssa da Veiga de. **Guia sobre a construção de tesouros**. Brasília: Ibict, 2021

SILVA, Armando Malheiro da. O método quadripolar e a pesquisa em Ciência da Informação. **Prisma.com (Portugal)**, n. 26, p. 27-44, 2014. Disponível em: <https://brapci.inf.br/index.php/res/download/86489>. Acesso em: 03 dez. 2022

SILVA, Armando Malheiro da; RIBEIRO, Fernanda. **Das “ciências” documentais à Ciência da Informação: ensaio epistemológico para um novo modelo curricular**. Porto: Ed. Afrontamento, 2002.

SILVA FILHO, Jomson Teixeira da. O signo linguístico entre Saussure e Benveniste: ainda de sua natureza “arbitrária” e “necessária” e sua relação com o sujeito falante/locutor. **Linguagem & Ensino**, v. 23, n. 3, p. 838-861, jul./set. 2020

SMIT, Johanna Wilhelmina. A busca pela eficiência na representação da informação e do conhecimento: desdobramentos posteriores no pensamento de Gardin. In: GUIMARÃES, J. A. C e Dodebei, V.; (orgs). **Organização do conhecimento e diversidade cultural**. Marília: ISKO-Brasil: FUNDEPE, 2015. 835 p. (Série: Estudos Avançados em Organização e Representação do Conhecimento, v. 3).

SMIT, Johanna Wilhelmina. A documentação e suas diversas abordagens. In: GRANATO, Marcus; SANTOS, Cláudia Penha dos; LOUREIRO, Maria Lucia de Niemeyer Matheus (Orgs.). **Documentação em museus**. Rio de Janeiro: MAST, 2008. p. 11-22.

SOUSA, Janailton Lopes. **Avaliação do padrão Simple Knowledge Organization System (SKOS) para a representação de vocabulários controlados**. Orientador: Rogério Aparecido Sá Ramalho. 2019. 100f. – Dissertação (Dissertação de Mestrado em Ciência da Informação) – Universidade Federal de São Carlos (UFSCar), Programa de Pós-Graduação em Ciência da Informação, São Carlos, 2019.

SOUSA, Janailton Lopes; MARTINS, Paulo George Miranda; RAMALHO, Rogério Aparecido Sá. Análise dos padrões XML e RDF para a representação na web sob a perspectiva da Ciência da Informação: um estudo preliminar. **Informação &**

**Tecnologia** (ITEC), v. 5, p. 65-78, 2018a. Disponível em: <https://periodicos.ufpb.br/index.php/itec/article/view/38458/22109>. Acesso em 13 jan. 2021

SOUSA, Janailton Lopes; MARTINS, Paulo George Miranda; RAMALHO, Rogério Aparecido Sá. Modelos de Representação Semântica na Era do Big Data. **Brazilian Journal of Information Science**, v. 12, p. 34-40, 2018b. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/8167/5335>. Acesso em: 12 jan. 2021

SOWA, John F. **Conceptual structures: information processing in mind and machine**. Includes indexes [S.l.]: Addison-Wesley; 1st Printing edition, 1983. 481 p. (The Systems programming series)

STELLATO, Armando; RAJBHANDARI, Sachit; TURBATI, Andrea; FIORELLI, Manuel; CARACCILOLO, Caterina; LORENZETTI, Tiziano; KEIZER, Johannes; PAZIENZA, Maria Teresa. Collaborative Development of Multilingual Thesauri with VocBench (System Description and Demonstrator). **ESWC 2015: The Semantic Web: ESWC 2015 Satellite Events**, 2015a. Disponível em: [https://link.springer.com/chapter/10.1007/978-3-319-25639-9\\_29](https://link.springer.com/chapter/10.1007/978-3-319-25639-9_29). Acesso em: 14 ago. 2022

STELLATO, Armando; RAJBHANDARI, Sachit; TURBATI, Andrea; FIORELLI, Manuel; CARACCILOLO, Caterina; LORENZETTI, Tiziano; KEIZER, Johannes; PAZIENZA, Maria Teresa. VocBench: a Web Application for Collaborative Development of Multilingual Thesauri. **ESWC 2015: The Semantic Web. Latest Advances and New Domains**, 2015b. Disponível em: [https://link.springer.com/chapter/10.1007/978-3-319-18818-8\\_3](https://link.springer.com/chapter/10.1007/978-3-319-18818-8_3). Acesso em: 14 ago. 2022

TÁLAMO, Maria de Fátima G. Moreira; LARA, Marilda Lopes Ginez de; KOBASHI, Nair Yumiko. Contribuição da terminologia para a elaboração de tesouros. **Ciência da Informação**, v. 21, n. 3, p. 197-200, set./dez. 1992

TÁLAMO, Maria de Fátima G. Moreira; SMIT, Johanna Wilhelmina. Information science: informational concept and disciplinary integration. **Brazilian Journal of Information Science**, v. 1, n. 1, 2007. DOI: 10.5016/brajis.v1i1.30 Acesso em: 14 mar. 2021.

TESSER, Gelson João. Principais linhas epistemológicas contemporâneas. **Educar em Revista**, v. 10, n. 10, p. 91-98, 1995. Disponível em: <https://revistas.ufpr.br/educar/article/view/36044/22233>. Acesso em: 10 mar. 2021

TOUTAIN, Lídia Maria Batista Brandão (org.). **Para entender a ciência da informação**. Salvador : EDUFBA, 2007. 242 p.

VALENTIM, Marta Lígia Pomim (Org.). **Métodos qualitativos de pesquisa em Ciência da Informação**. São Paulo: Polis, 2005.

VECHIATO, F. L. **Encontrabilidade da informação: contributo para uma conceituação no campo da ciência da informação**. 2013. 206 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013.

WILL, Leonard. The ISO 25964 Data Model for the Structure of an Information Retrieval Thesaurus. **Bulletin of the American Society for Information Science & Technology**, v. 38, n. 4, p. 48-51, 2012

ZACARIAS, Rodrigo Oliveira; JACYNTHO, Mark Douglas de Azevedo; VASCONCELOS, Aline Pires Vieira de. Um método para criação de tesauros semânticos de requisitos usando a ontologia SKOS. **Transinformação**, v. 32, e190086, 2020. Disponível em: <https://doi.org/10.1590/2318-0889202032e190086>. Acesso em: 18 out. 2021

ZENG, Marcia Lei; MAYR, Philipp. Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review. **International Journal on Digital Libraries**, v. 20, p. 209–230, 2018