

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

JANAILTON LOPES SOUSA

**AVALIAÇÃO DO PADRÃO SIMPLE KNOWLEDGE ORGANIZATION SYSTEM
(SKOS) PARA A REPRESENTAÇÃO DE VOCABULÁRIOS CONTROLADOS**

São Carlos
2019

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

JANAÍLTON LOPES SOUSA

**AVALIAÇÃO DO PADRÃO SIMPLE KNOWLEDGE ORGANIZATION SYSTEM
(SKOS) PARA A REPRESENTAÇÃO DE VOCABULÁRIOS CONTROLADOS**

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

Orientador: Prof. Dr. Rogério Aparecido Sá Ramalho.

Apoio: FAPESP – Processo 2017/21121-1

São Carlos
2019

Sousa, Janailton Lopes

Avaliação do padrão Simple Knowledge Organization System (SKOS) para a representação de vocabulários controlados / Janailton Lopes Sousa. -- 2019. 100 f. : 30 cm.

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

Orientador: Rogério Aparecido Sá Ramalho

Banca examinadora: Rogério Aparecido Sá Ramalho, Zaira Regina Zafalon, Brígida Maria Nogueira Cervantes, Cibele A. C. Marques dos Santos Bibliografia

1. KOS. 2. Vocabulários controlados. 3. SKOS. I. Orientador. II.

Universidade Federal de São Carlos. III. Título.

Ficha catalográfica elaborada pelo Programa de Geração Automática da Secretaria Geral de Informática (SIn).

DADOS FORNECIDOS PELO(A) AUTOR(A)

Bibliotecário(a) Responsável: Ronildo Santos Prado – CRB/8 7325

ERRATA

Folha	Onde se lê	Leia-se
1	Apoio: FAPESP – Processo 2017/21121-1	Apoio: Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP – Processo 2017/21121-1 e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES
6	À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo suporte, credibilidade e confiabilidade; à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio ao PPGCI/UFSCar.	À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) – processo 2017/21121-1, obrigado pelo suporte, credibilidade e confiança; à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio ao PPGCI/UFSCar.

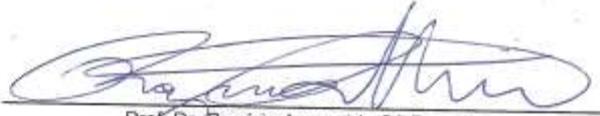


UNIVERSIDADE FEDERAL DE SÃO CARLOS

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

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Janailton Lopes Sousa, realizada em 13/08/2019:



Prof. Dr. Rogério Aparecido Sá Ramalho
UFSCar

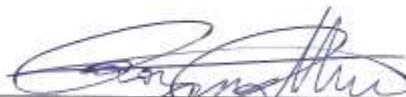
Profa. Dra. Zaira Regina Zafalon
UFSCar

Profa. Dra. Brígida Maria Nogueira Cervantes
UEL

Profa. Dra. Cibele Araújo Camargo Marques dos Santos
USP

~~Assinatura do Candidato~~

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Zaira Regina Zafalon, Brígida Maria Nogueira Cervantes, Cibele Araújo Camargo Marques dos Santos e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ão) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.



Prof. Dr. Rogério Aparecido Sá Ramalho

DEDICATÓRIA

Dedico à minha família e especialmente a minha amada Joyce pela coragem em aceitar os desafios que lhe foram propostos para chegar até aqui. A todos que contribuíram para concretização desta etapa.

AGRADECIMENTOS

A Deus, cujo amor e misericórdia superam o entendimento humano.

Agradeço à minha família, que abdicou do desejo de estar comigo em favor da realização desta etapa de minha vida.

À minha amada Joyce, que de mãos dadas resolveu enfrentar este desafio e em todos os momentos esteve comigo, seu companheirismo é mister.

Aos professores do Curso de Biblioteconomia da Universidade Federal do Maranhão, cujo apoio foi essencial aos primeiros passos que galguei na pesquisa científica até o ingresso no mestrado.

Ao meu orientador Rogério Ramalho, cujo suporte foi vital para a continuação de minha caminhada acadêmica, pois, com sabedoria, humildade e sensibilização ao próximo, transmitiu ensinamentos valorosos.

A todos os professores do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos - (PPGCI/UFSCar), pelo apoio e participação no processo de construção e desconstrução das propriedades conceituais.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo suporte, credibilidade e confiabilidade; à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio ao PPGCI/UFSCar.

Aos meus amigos Paulo, Sandra, Miguel e Cris, cuja empatia sobrenatural me auxiliou desde o começo desta etapa, expressei minha sincera gratidão. Aos meus amigos e colegas, Aderlou, Josy, Cantanheide, Raimunda Marinho, Valdirene Conceição, Marcos Aurélio, Jhonatan Jardim, Valéria e Glória Alencar, cujo apoio me deu condições para pleitear a vaga de ingresso no mestrado, obrigado por acreditarem em mim.

Aos colegas da 2ª turma do mestrado, pelas discussões e compartilhamento de experiências e conhecimentos.

A todos aqueles que direta ou indiretamente contribuíram para a realização desta etapa.

RESUMO

SOUSA, Janailton Lopes. **Avaliação do padrão Simple Knowledge Organization System (SKOS) para a representação de vocabulários controlados**. 2019. 100 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de São Carlos, São Carlos, 2019.

Knowledge Organization System (KOS) são instrumentos de representação de recursos informacionais utilizados na Ciência da Informação, o *background* do KOS inclui o estudo dos conceitos e uso de instrumentos tecnológicos digitais. O *Simple Knowledge Organization System* (SKOS) representa essa junção entre a fundamentação teórica do KOS e suas estruturas conceituais, com o uso de ferramentas digitais de representação, favorecendo a representação de um KOS tradicional em ambientes digitais. Todavia, o aumento nas publicações de vocabulários controlados, disponibilizados em ambientes digitais, acena para a necessidade de identificação de novos procedimentos de avaliação desses instrumentos. Neste sentido, o objetivo geral desta pesquisa é identificar parâmetros de avaliação do padrão SKOS, para a representação de vocabulários controlados. Esta pesquisa é de cunho teórico e metodológico, de abordagem qualitativa e caráter exploratório, que se propõe a examinar os temas de representação, organização, tecnologias semânticas e disponibilização de vocabulários em ambientes digitais. Como resultado desta pesquisa, são apresentados parâmetros de avaliação que podem ser aplicados em vocabulários codificados em SKOS, servindo, portanto, como fonte referencial que pode auxiliar pesquisas nesta temática. Conclui-se que a dificuldade de avaliar vocabulários controlados ainda é presente, neste sentido, o uso de parâmetros que possam auxiliar este processo fornece um contributo no avanço das discussões que incluem esta temática no campo de estudo da Organização do Conhecimento.

Palavras-chave: Knowledge Organization System. Vocabulários controlados. Tesouros. Simple Knowledge Organization System.

ABSTRACT

SOUSA, Janailton Lopes. **Evaluation of the Simple Knowledge Organization System (SKOS) Standard for Controlled Vocabulary Representation**. 2019. 100 f. Dissertation (Master in Information Science) - Federal University of São Carlos, São Carlos, 2019.

Knowledge Organization System (KOS) are information resources representation tools used in Information Science, the KOS background includes the study of the concepts and use of digital technological instruments. The Simple Knowledge Organization System (SKOS) represents this junction between the KOS theoretical foundation and its conceptual frameworks, with the use of digital representation tools, favoring the representation of a traditional KOS in digital environments. However, the increase in controlled vocabulary publications, available in digital environments, points to the need to identify new evaluation procedures for these instruments. In this sense, the general objective of this research is to identify evaluation parameters of the SKOS standard, for the representation of controlled vocabularies. This research is theoretical and methodological, with a qualitative approach and exploratory character, which aims to examine the themes of representation, organization, semantic technologies and vocabulary availability in digital environments. As a result of this research, evaluation parameters are presented that can be applied in SKOS-coded vocabularies, thus serving as a reference source that can assist research in this theme. It is concluded that the difficulty of evaluating controlled vocabularies is still present, in this sense, the use of parameters that can help this process contributes to the advancement of discussions that include this theme in the field of study of the Knowledge Organization.

Keywords: Knowledge Organization System. Vocabularies controlled. Thesauri. Simple Knowledge Organization System.

LISTA DE FIGURAS

Figura 1 – Esferas de conceitos de Schopenhauer.....	26
Figura 2 – Variedade de KOS.....	41
Figura 3 – Descrição simples de livro em XML.....	49
Figura 4 – Schema XML para validação.....	49
Figura 5 – Tripla RDF	50
Figura 6 – RDF com prefixos DC.....	51
Figura 7 – Propriedades do skos:broader.....	59
Figura 8 – Propriedades do skos:broaderTransitive.....	59
Figura 9 – Modelo de vocabulário em SKOS.....	60
Figura 10 – Áreas do conhecimento TBCI.....	68
Figura 11 – Trecho das áreas do conhecimento do TBCI em SKOS.....	69
Figura 12 – Representação gráfica da conversão do TBCI em SKOS.....	70

LISTA DE QUADROS

Quadro 1 – KOS- Vocabulários e suas definições	40
Quadro 2 – Terminologia para Vocabulários Controlados.....	43
Quadro 3 – Vocabulário SKOS.....	55
Quadro 4 – SKOS Note e duas definições.....	57
Quadro 5 – Tipos de avaliação de vocabulários.....	73
Quadro 6 – Categorização dos erros identificados.....	80
Quadro 7 – Definição dos erros.....	81

LISTA DE ABREVIATURAS E SIGLAS

ABNT Associação Brasileira de Normas Técnicas
ANSI American National Standard Institute
CDD Classificação Decimal de Dewey
CDU Classificação Decimal Universal
HTML HyperText Markup Language
IBICT Instituto Brasileiro de Informação em Ciência e Tecnologia
IRI Identificador de Recursos Internacionalizados
ISKO International Society for Knowledge Organization
ISO International Organization for Standardization
KO Knowledge Organization
KOS Knowledge Organization System
MIT Massachusetts Institute of Technology
NISO National Information Standards Organization
OC Organização do Conhecimento
OWL Web Ontology Language
RDF Resource Description Framework
SGML Standard Generalized Markup Language
SKOS Simple Knowledge Organization System
SOC Sistema de Organização do Conhecimento
SRI Sistema de Recuperação de Informação
TCT Teoria Comunicativa da Terminologia
TGT Teoria Geral da Terminologia
UFSCar Universidade Federal de São Carlos
URI Uniform Resource Identifier
W3C World Wide Web Consortium
WWW World Wide Web
XML eXtensible Markup Language

SUMÁRIO

1 INTRODUÇÃO	13
1.1 Problema de pesquisa	18
1.2 Justificativa	18
1.3 Objetivos	19
1.3.1 Objetivo geral.....	19
1.3.2 Objetivos específicos.....	19
1.4 Procedimentos metodológicos	19
2 ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO	22
2.1 Aspectos filosóficos da representação do conhecimento	23
2.2 Knowledge Organization Systems	35
2.3 Tecnologias Semânticas	46
3 SIMPLE KNOWLEDGE ORGANIZATION SYSTEM	53
3.1 Vocabulários controlados em SKOS	63
3.2 Parâmetros para avaliação do padrão SKOS	71
4 CONSIDERAÇÕES FINAIS	83
REFERÊNCIAS	88
APÊNDICE	98

1 INTRODUÇÃO

A Ciência da Informação surge de forma integrada com outras áreas do conhecimento, como a Biblioteconomia, Ciência da Computação, Comunicação, Linguística, entre outras. Para Pinheiro e Loureiro (1995), na década de 1960 são elaborados os primeiros registros teóricos da nova área, o que implica uma tentativa de melhor demarcá-la e estabelecer relações interdisciplinares com outros campos de estudos, cujo objeto de pesquisa é a informação.

Os métodos de transmissão e processamento da informação não foram desenvolvidos instantaneamente, apesar do constante surgimento de novas tecnologias que otimizam a sistematização de informações. A informação e o conhecimento em sociedade superam o simples estereótipo tecnológico de seu gerenciamento, envolvem novas formas de organização social, econômica e política. Isto porque a informação transporta algo de imaterial, que permeia as diversas áreas da sociedade, portanto, existe uma espécie de Revolução Informacional (LOJKINE, 1995), que também é social, além de tecnológica.

As razões para o surgimento da Ciência da Informação (CI) são de cunho militar e tecnológico, especificamente nos Estados Unidos, entretanto, a Documentação é uma importante propositora de métodos para a sistematização do que é entendido por informação e conhecimento. Oliveira (2005) deixa claro seu posicionamento sobre as contribuições da Documentação para a origem da CI, principalmente pelo surgimento do Instituto Internacional de Bibliografia, dos conceitos de documento, bibliografia como registro, memória do conhecimento científico e a Classificação Decimal Universal. Silva e Freire (2012) também enfatizam as raízes da identidade da CI e destacam a contribuição da Biblioteconomia e da Documentação.

A Documentação foi um importante expoente para alavancar o entendimento sobre informação, bem como promover ideias inovadoras sobre tecnologia aplicada aos centros de informação e suportes documentais (SILVA; FREIRE, 2012). O relacionamento das contribuições de Naudé, Dewey e Otlet, é apresentado sob uma ótica temporal de acordo com estudos de Tálamo e Smit (2007), também reforçados por Silva e Freire (2012). Esta visão precursora da CI, penetra diretamente na sua formação epistemológica, onde os estudos da Computação, Linguística, Comunicação, etc., são identificados como áreas correlatas ou componentes formativos desta ciência.

O nascimento da Ciência da Informação é envolvido por um período que se estende do final da segunda Guerra Mundial, as invenções elétricas e mecânicas fizeram parte desse contexto, como sugere a publicação do artigo *As We May Think* no *Atlantic Monthly* em 1945 (BUSH, 1945), pelo então cientista do *Massachusetts Institute of Technology* – (USA), Vannevar Bush (OLIVEIRA, 2005). A proposição do artigo de Bush se direcionava à organização e armazenamento das informações e dados militares em constante crescimento, com o auxílio de uma máquina intitulada *MEMEX*, acrônimo de *Memory Extensible* ou memória extensível.

Esse espaço temporal subsidia os fundamentos tecnológicos, que norteiam os processos de representação com o uso de tecnologias eletrônicas e estabelecem um novo rumo no tratamento sistemático da informação. De acordo com Silva e Freire (2012), o advento da Ciência da Informação parte de um arrazoado técnico-científico carente de objetivos, finalidades, métodos, além de noção escassa de tecnologia para construção da identidade da área. Ou seja, a CI colocou à margem algumas deficiências teóricas e pragmáticas, que surgiram no pós-guerra, por isso a visão tecnicista americana se opunha às proposições humanistas advindas da Europa.

Neste aspecto, a informação e o conhecimento gerenciados com o suporte teórico da Documentação e da Biblioteconomia, entram em conflito terminológico, documento/informação, enquanto o objeto da Ciência da Informação se apresenta, fluido e intangível, obliterando as concepções que tentam defini-lo em sua ligação inexorável com a tecnologia de informação (SARACEVIC, 1996). Isto foi motivo para usar a Teoria Matemática da Comunicação de Shannon e Weaver (1949), da Universidade de Illinois, como o tabernáculo desta ciência, que emerge para resolver problemas relacionados à informação com uso de tecnologias eletrônicas. Birger Hjørland (1998), da Escola Real de Biblioteca e Ciência da Informação em Copenhague, na Dinamarca, oferece uma crítica muito pertinente a respeito das teorias e meta-teorias da Ciência da Informação, ou a sua falta.

É um fato bem conhecido que a CI carece de boas teorias, a maioria dos trabalhos são pragmáticos, que resistem à análise científica e à generalização. No entanto, muitos dos trabalhos são publicados e muito trabalho prático é feito sem explicar nenhum pressupostos teóricos ou meta-teóricos. Isso torna muito difícil fazer um trabalho histórico e filosófico em CI. No entanto, por trás de todos os tipos de atividades são certas suposições sobre o mundo em geral, sobre humanos seres, sobre linguagem, cognição, pesquisa, e assim por diante. (HJØRLAND, 1998, p.607, tradução nossa).

Neste ponto Hjørland (1998), volta-se para a necessidade de maior reflexão na Ciência da Informação, do ponto de vista da construção de suas teorias, que ora ou outra beneficia-se de teorias de outras áreas para justificar-se. A crítica de Hjørland (1998) se estende

até a necessidade do uso da filosofia na CI, que, lembrando o pensamento de Bachelard (2006), ressalta a necessidade de “acordar de seu sono” para fortalecer a si mesmo e as ciências modernas.

A filosofia é o campo do conhecimento, onde a maioria dos conhecimentos gerais desse tipo é armazenado e organizado. A filosofia aprende das ciências singulares, generaliza esse conhecimento e comunica-se com esse conhecimento de volta em diferentes Ciências em diferentes quantidades. A CI pode aprender com a filosofia, mas a filosofia não pode ditar princípios a outras ciências. Deve haver cooperação, e a CI deve compreender seus próprios problemas filosóficos (HJØRLAND, 1998, p.607, tradução nossa).

Neste sentido, a necessidade de uma introspecção teórica, com auxílio da Filosofia, se mostra presente na CI, para compreensão de seus fenômenos epistemológicos. Para emular esta simbiose entre Filosofia e tecnologia na CI, são apresentadas algumas reflexões iniciais sobre representação, conceito e linguagem, que ao longo do tempo se incorporaram nos *Knowledge Organization Systems* (KOS) e que muitas vezes passam despercebidas, devido ao amortecimento que a tecnologia causa nas atividades diárias. As interseções entre os fundamentos basilares aplicados às novas tecnologias digitais têm demonstrado uma eficiente colaboração no processo de representação do conhecimento.

O contexto que situa as discussões na Ciência da Informação é marcado pela influência tecnológica. Todavia, o uso de técnicas que subsidia as práticas do profissional da informação é destacado pela consistência e perenidade, que os modelos tradicionais de representação da informação fixaram nas unidades de informação. O desafio, que desencadeia o cerne da prática profissional, é impulsionado pela dinâmica das transformações sociais influenciadas pelo uso de tecnologias digitais.

As novas tecnologias digitais possibilitaram um avanço considerável no campo da Ciência da Informação ao longo das últimas décadas, favorecendo o desenvolvimento de novos instrumentos e métodos de representação, armazenamento, organização, processamento e recuperação de informações. Conforme relata Lima (2015, p.683-684):

Este avanço impactou primeiramente nas atividades relacionadas ao processo de indexação, com o surgimento de sistemas automatizados para criação de índices e processamento automático do texto, com os aportes da linguística, no processo de catalogação, com a criação do formato MARC, pela *Library of Congress* (LC), que passa a ser utilizado efetivamente em 1966, e que foi considerado pela *International Standard Organization* (ISO) o primeiro projeto de automação internacional a estabelecer normas de descrição bibliográfica em forma mecanicamente legível, possibilitando o intercâmbio de dados.

Do mesmo modo, por meio da incorporação das tecnologias computacionais nos processos de classificação, tornou-se possível a compatibilização de vocabulários controlados com auxílio do computador, favorecendo a integração de tesouros em Sistemas de Recuperação de Informação (SRI) (ALVITE DÍEZ et al., 2010). Nesta perspectiva, observa-se que as características inerentes aos novos ambientes digitais, exigem o desenvolvimento de outros tipos de habilidades e aperfeiçoamento das técnicas profissionais tradicionalmente utilizadas no campo da Ciência da Informação.

As facilidades de intercâmbio, graças às tecnologias de informação vigentes, passam a requerer maior atenção dos profissionais de informação ao planejarem seus serviços, levando em conta os padrões ou em desenvolvimento ou já instituídos para as diferentes atividades que ocorrem no interior de um serviço de recuperação de informação (GOMES; CAMPOS, 2004, p.2).

Ao longo dos últimos anos, inúmeras iniciativas têm sido propostas no intuito de desenvolver e aperfeiçoar métodos e ferramentas que possam potencializar o uso de linguagens e vocabulários controlados em ambientes computacionais, buscando favorecer um melhor atendimento das demandas informacionais contemporâneas, originando um crescente número de pesquisas relacionadas ao desenvolvimento de KOS.

Segundo Hodge (2000), o termo *Knowledge Organization Systems* (KOS) foi proposto em 1998, pelo *Networked Knowledge Organization Systems Working Group*, para englobar sistemas de classificação, cabeçalhos de assunto, arquivos de autoridade, redes semânticas e ontologias. Na classificação de tipos de KOS, proposta por Hodge (2000), são incluídos dicionários, glossários, taxonomias e tesouros. Conforme destaca Lara (2015, p. 92), “[...] no Brasil, não há consenso sobre a utilização de um termo que abrange o conjunto de instrumentos de organização da informação e do conhecimento”. Apesar das variações de denominação identificadas na literatura nacional para designar tais instrumentos, observa-se um crescente aumento no número de pesquisas relacionadas à essa temática.

A Teoria do Conceito tem se destacado como aporte teórico no âmbito da Organização do Conhecimento, devido às características que tornam o conceito como unidade do conhecimento. No entanto, o conceito não se limita a essa definição, sua exploração no âmbito filosófico, por exemplo, explicita sua intrínseca relação com a Lógica, a partir disso é possível perceber que os conceitos podem ser resultados de cálculos lógicos, dentre outras características que podem ser identificadas em KOS.

A característica básica de um KOS reside no fato de elencar termos que possam representar conceitos. Neste aspecto, os termos não devem ser resultados apenas de uma seleção

temática, pois, levando em consideração essa característica conceitual, Schopenhauer (2005) e Dahlberg (1978) destacam que os conceitos se aplicam às leis da Lógica, portanto, seu uso precisa estar alinhado às premissas do raciocínio lógico. A exploração dessas nuances do conceito demonstra sua grande versatilidade, que parte de um ponto abstrato até sua instrumentalização por meio de KOS.

Os KOS são estruturas resultantes dos arranjos entre conceitos, logo, as percepções teóricas, que caracterizam a noção de representação no sentido filosófico, são apresentadas como *background* das tecnologias que envolvem os vocabulários controlados. O conceito é um elemento que mantém constante suas propriedades, pois, no ambiente digital é caracterizado como um identificador único, ou seja, é uma unidade indivisível. O relacionamento entre esses conceitos, no ambiente digital, é expresso por meio de triplas semânticas que exprimem o contexto dos recursos informacionais por meio do *Resource Description Framework* (RDF).

A partir da popularização do ambiente Web, novos instrumentos de organização e representação têm sido propostos, provocando convergências entre aportes teóricos e metodológicos de diversas campos científicos, no intuito de atender as demandas informacionais de uma Sociedade cada vez mais baseada em interações virtuais (RAMALHO, 2010). A área de Ciência da Informação sempre preocupou-se com aspectos semânticos da representação, pois, conforme destaca Le Coadic (2004), é inerente à toda informação um elemento de sentido; em relação ao enfoque das representações semânticas, Zafalon (2013) esclarece que é a semântica que dá conta dos processos mentais segundo os quais se produz, constitui, compreende e descreve a representação de um recurso informacional. Ou seja, a semântica é parte inerente do processo de representação.

No ano de 2009, o *World Wide Web Consortium* (W3C) formalizou, como Recomendação, o padrão *Simple Knowledge Organization System* – SKOS (Sistema de Organização do Conhecimento Simples), como um modelo de dados para a representação de KOS no ambiente Web, buscando atender às novas demandas informacionais relacionadas ao compartilhamento e interoperabilidade de vocabulários no ambiente Web (RAMALHO, 2015b).

O SKOS fornece um modelo para a representação formal de Sistemas de Organização: como tesouros, esquemas de classificação, listas de cabeçalho de assunto, taxonomias, entre outros tipos similares de vocabulários controlados (RAMALHO; VIDOTTI; FUJITA, 2007). Sua codificação é baseada no padrão RDF, favorecendo a representação de

informações de maneira estruturada, assim como a integração de diferentes esquemas conceituais.

Desde a formalização do SKOS, a publicação de vocabulários controlados em ambientes digitais tem aumentado, não que este seja a única causa desse aumento, mas não se pode descartar sua contribuição nesse processo. Nayak (2019) explica que esses vocabulários têm evidenciado a necessidade de criar meios inteligentes de avaliação, pois, vocabulários publicados na forma de dados vinculados, possuem centenas de termos e relacionamentos, com possibilidade de expansão constante.

Somado a isso, a avaliação de vocabulários controlados não tem se destacado tanto, como outras temáticas exploradas na OC, por isso, há poucos registros na literatura de métodos que podem ser utilizados. Portanto, há uma necessidade de identificar os fundamentos teóricos que estão diretamente relacionados com os tipos de KOS e padrões utilizados para representar vocabulários controlados, assim como possíveis meios de avaliá-los.

Destaca-se a relevância desta pesquisa considerando que o SKOS pleiteia o surgimento de um novo campo de investigação eminentemente interdisciplinar, capaz de se engendrar na constelação epistemológica no campo da Organização e Representação do Conhecimento (ORC), apontando novas possibilidades de discussão na área da Ciência da Informação.

1.1 Problema de pesquisa

Quais parâmetros são empregados no processo de avaliação do SKOS para a representação de vocabulários controlados.

1.2 Justificativa

Os KOS são referências macros para a construção de vocabulários, as áreas gênicas da Ciência da Informação são as precursoras desse instrumento, que ao longo do tempo foi influenciado pelo uso de tecnologias digitais, em contrapartida às normas que orientam o processo de elaboração de vocabulários que têm acompanhado essa influência e se adaptado a essas transformações, como foi o caso da ISO 25964 (2011) e (2013).

Apesar desse notável avanço na evolução dos vocabulários publicados em ambientes digitais, a NATIONAL INFORMATION STANDARDS ORGANIZATION (2010), desde 2005 ressalta que há mínima preocupação quando o assunto é avaliação de vocabulários. O que evidencia a necessidade de estudos que explorem os meios de avaliação disponíveis e como isso implica no uso dos vocabulários, assim como o uso de técnicas que possam auxiliar

nesse processo. A inserção do SKOS como aplicação do RDF, que garante semântica por meio do relacionamento entre conceitos, demonstra a potencialidade dos vocabulários no ambiente digital, porém destaca uma preocupação com sua qualidade.

Os resultados, obtidos a partir da revisão de literatura realizada, possibilitam o desenvolvimento de uma reflexão teórica para a discussão dos diferentes pontos de vista sobre o tema proposto, favorecendo, assim, a criação de subsídios para um melhor entendimento sobre a utilização do SKOS. A partir dos estudos realizados, serão apresentados os principais parâmetros de avaliação identificadas ao longo da pesquisa, no intuito de contribuir para a divulgação e popularização de novos métodos de representação de vocabulários em ambientes digitais.

A avaliação de vocabulários em SKOS, publicados no ambiente digital, reflete diretamente na integridade e disponibilidade de recursos informacionais representados na web. Logo, a relevância e o impacto social desta pesquisa, pauta-se na verificação dos *status* de qualidade de vocabulários em SKOS. O que contribui para melhores práticas para publicar vocabulários vinculados, assim como sua manutenção.

1.3 OBJETIVOS

1.3.1 Objetivo geral

Identificar parâmetros de avaliação do padrão SKOS para a representação de vocabulários controlados.

1.3.2 Objetivos Específicos

Identificar, na literatura científica, pesquisas, experiências, projetos e normas internacionais que abordam a temática de publicação de vocabulários em SKOS;

Examinar os fundamentos que norteiam SKOS, analisando aproximações e divergências em relação aos instrumentos de representação tradicionalmente utilizados no campo da Ciência da Informação;

Apresentar parâmetros de avaliação para vocabulários em SKOS.

1.4 Procedimentos metodológicos

A definição de uma metodologia é um procedimento padrão na comunidade científica, pois indica o caminho a ser percorrido durante a pesquisa. Por meio da metodologia é possível estabelecer um modelo conceitual e prático a ser seguido, que certamente é

reconhecido pela academia científica para estudar fenômenos, sujeitos, propriedades e suas relações.

Para González de Gómez (2000), a metodologia da pesquisa designa, de maneira ampla, o início e orientação de um movimento de pensamento, cujo esforço e intenção direcionasse à produção de um novo conhecimento. Neste sentido, a metodologia consiste em estabelecer parâmetros que referenciam o trabalho do pesquisador e garante a perenidade do conhecimento produzido com base nos cânones do pensamento científico.

Trata-se de uma pesquisa de cunho teórico e metodológico, que se propõe estudar os temas de representação, organização e disponibilização de vocabulários em ambientes digitais. Possui abordagem qualitativa, pois busca apresentar características voltadas à descrição, compreensão e explicação de relações de um determinado fenômeno, de modo a considerar o caráter interativo entre os objetivos definidos e as orientações teóricas da pesquisa.

O caráter exploratório da pesquisa, segundo Cervo e Bervian (2003), consiste na busca de informações sobre o assunto a ser estudado, considerando os diversos aspectos de um problema, a fim de solucioná-lo. Busca tornar o problema mais explícito a partir de hipóteses, o que exige descrição de fatos observados (GIL, 2008; TRIVIÑOS, 1987). Segundo Köche (2002, p. 126), o estudo exploratório refere-se a “[...] um processo de investigação que identifique a natureza do fenômeno e aponte as características essenciais das variáveis que se quer estudar”. Além disso, o caráter descritivo da pesquisa permitirá descrever as características do objeto de estudo investigado (COSTA, COSTA, 2001).

A análise da temática é de caráter exploratório da literatura disponível sobre o tema, para oportunizar uma pesquisa de caráter analítico e descritivo das características sobre o objeto em estudo, apresentado nos objetivos propostos. A partir de uma análise da literatura, é possível construir um conhecimento teórico sobre o tema investigado, a fim de auxiliar a construção e compreensão do problema e do objeto de pesquisa.

Partindo da necessidade de compreender os fundamentos filosóficos da representação, assim como as normas que orientam a publicação de vocabulários e as tecnologias digitais que fazem parte desse processo. Esta pesquisa explora as principais noções de representação, linguagem e conceito para subsidiar teoricamente os instrumentos de representação utilizados na Organização Representação do Conhecimento.

À medida que se explora tais fundamentos teóricos, é possível encontrar determinados conceitos que estão intrinsecamente relacionados com os instrumentos de

representação, e mesmo evoluindo para o ambiente digital, mantém suas características identificáveis. Quanto aos aspectos metodológicos, as recomendações SKOS, publicadas pela W3C, são importantes fontes de pesquisa que nortearão esta pesquisa.

Este trabalho está dividido em quatro partes, a primeira apresenta a introdução, com a questão da pesquisa, da justificativa, dos objetivos e procedimentos metodológicos. Nessa parte são previamente resumidas algumas perspectivas exploradas ao longo do trabalho. Na segunda parte há uma discussão sobre Organização e Representação do Conhecimento (ORC), destacando os aspectos filosóficos da representação, as propriedades dos KOS e o das tecnologias semânticas.

Na terceira parte é apresentado o padrão SKOS com suas principais características, além de vocabulários controlados em SKOS e os parâmetros que podem ser utilizados para avaliar vocabulários controlados em SKOS. Na quarta e última parte, são apresentadas as considerações finais da pesquisa, com algumas reflexões sobre a importância da revisão teórica conceitual que envolve a OC e sobre a avaliação dos instrumentos de representação utilizados na Ciência da Informação.

2 ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

O conhecimento sobre os objetos, ações e fenômenos é resultado de diversas percepções cognoscíveis do ser humano, por meio do espaço e do tempo, nisto é possível elencar atributos que possam representá-lo em determinados domínios. Para tanto, é natural recorrer aos métodos de organização e representação do conhecimento para sistematizar estes atributos, que passam por um tratamento técnico amparado pelas teorias da Biblioteconomia e Ciência da Informação.

A representação foi uma das formas encontradas pelo ser humano para submeter o conhecimento às condições sobre as quais estão regidas a realidade, quer dizer, o tempo e o espaço. Embora sua origem esteja relacionada a essas condições, no mundo das ideias torna-se apenas uma referência às percepções externas. Logo, foi necessário encontrar formas para representar e racionalizar sua existência. Neste sentido, os instrumentos elaborados ao longo do tempo para organizá-lo, são alternativas encontradas para este processo de transmutação do abstrato para o concreto.

A Organização do Conhecimento (OC) tem contribuído incessantemente para o avanço da Ciência da informação, a *International Society for Knowledge Organization* (ISKO) é a prova disso, criada em julho de 1989, na Alemanha, por Ingetraut Dahlberg, agrega uma tríade temática composta pelas dimensões epistemológica, aplicada e cultural. De acordo com Guimarães (2015), a dimensão epistemológica volta-se para a consolidação de teorias, metodologias, paradigmas, escolas de pensamento. A dimensão aplicada está ligada ao avanço tecnológico e ao desenvolvimento de novos instrumentos. Já a dimensão cultural é voltada para os contextos sociais, as comunidades e os sujeitos que interagem na OC.

Este escopo teórico integrado é um dos alicerces que orientam as discussões sobre tecnologias que englobam os KOS, relacionados intrinsecamente com fundamentos filosóficos, linguísticos e tecnológicos. No âmbito filosófico, a principal discussão surge em torno do que é representação e conhecimento, de acordo com Schopenhauer (2005), o conhecimento é a relação entre o sujeito e o objeto. No âmbito linguístico, é por meio de um conjunto de termos significativos que se torna possível a representação. No âmbito tecnológico, refere-se diretamente à disposição das representações por meio de sua infraestrutura, quer seja por instrumentos eletrônicos ou não.

Esta intersecção entre os aspectos epistêmicos e aplicações da representação, tem revelado a perenidade que a teoria garante no desenvolvimento de novas tecnologias, logo, a

concepção de representação do conhecimento não pode sucumbir à uma visão limitada ou emblemática. Mas considerar a amplitude e complexidade do conhecimento humano sobre diferentes perspectivas, portanto, expõe-se a necessidade de discutir desde o aspecto mais denso, filosófico e abstrato ao mais efêmero, simples e concreto, para descobrir as conexões que encadeiam teoria e prática.

Dividido em três seções, este capítulo inicia com uma abordagem estritamente teórica sobre os aspectos filosóficos da representação do conhecimento, na **Seção 2.1**, o conceito é destacado como o elemento central dessa discussão, ressaltando a contribuição de Schopenhauer (2005), Dahlberg (1978) e Wittgenstein (1968) com algumas reflexões envolvendo representação, conhecimento, conceito e linguagem.

Posteriormente, há uma discussão sobre os KOS na **Seção 2.2**, com apresentação dos mecanismos que utilizam essa base teórica para viabilizar sua aplicação em instrumentos de representação do conhecimento, como vocabulários controlados, cuja evolução recebe influência de novas tecnologias. Apesar de perceber uma cisão entre a teoria e a prática, por se encontrarem em esferas distintas, será possível notar, ao longo deste capítulo, que as tecnologias aplicadas são apenas a materialização do amadurecimento teórico dos conceitos proeminentes nos diversos campos científicos, especialmente da Ciência da Informação e áreas afins.

Na **Seção 2.3**, são apresentadas as linguagens de marcação que representam a base das tecnologias semânticas, instrumentos que se configuram como *background* de vocabulários publicados em SKOS. Neste aspecto, são elencadas as características básicas da *eXtensible Markup Language* (XML) e do *Resource Description Framework* (RDF), que contribuem diretamente no processo de representação de vocabulários em SKOS no contexto da Ciência da Informação.

2.1 Aspectos filosóficos da representação do conhecimento

O conceito é amplamente discutido na literatura científica da Ciência da Informação, embora seja atribuída à Dahlberg (1978) a formulação da teoria do conceito na área, o filósofo alemão Arthur Schopenhauer antecede essa teoria e estabelece alguns princípios apresentados na sua obra *O Mundo como vontade e como representação*, a noção de conceito, neste trabalho essas noções são exploradas em consonância com a visão de Dahlberg (1978). Esta discussão substancia-se na visão filosófica de conceito e representação utilizada nos KOS, tendo em vista que o conceito é sua unidade básica de representação.

O conhecimento humano fundamenta-se em três leis ou condições, a saber: tempo, espaço e causalidade. Em outros termos, estas três leis ou condições são dispositivos da

inteligência humana, a concepção das coisas que nos chegam em imagens, condicionada por eles, se chama, pois, imanente; transcendente, assim atingiríamos a razão, depois de ter penetrado o caráter de meros modos de conhecimento que essas três formas interpostas têm (MANN, 1955).

Esta concepção de imagens relacionadas ao conhecimento, recupera a proposição de George Berkeley sobre noção da ideia imagem, onde explica que as nossas abstrações são resultados das formas de se perceber os objetos e fenômenos, com isso nossas impressões seriam formadas por uma imagem perceptiva que estaria ligada a uma determinada ideia (BERKELEY, 2010). Logo, o conceito é resultado dessas percepções que se encontram no universo abstrato, deste modo, a organização do conhecimento visa à construção de modelos de mundo que se constituem em abstrações da realidade (BRÄSCHER; CAFÉ, 2008).

O alto nível de abstração não elimina a possibilidade de materializar um conceito; um exemplo são as figuras geométricas, como triângulos, quadrados, retângulos, losangos, entre outras. Essas figuras geométricas não existem de fato na natureza, elas são resultado de um conglomerado de cálculos que definem sua forma, por meio da matemática; desse modo, é possível provar sua existência. Os conceitos se comportam deste modo, pois armazenam sua essência na abstração, porém nada impede que sejam convertidos em objetos físicos, processos ou termos significativos.

De acordo com Schopenhauer (2005), a essência completa dos conceitos, ou da classe de representações abstratas, reside exclusivamente na relação cujo princípio de razão exprime nelas; ora, como tal relação é o fundamento do conhecimento, segue-se que a representação abstrata possui sua essência inteira e, exclusivamente, em relação a outra representação que é o seu fundamento do conhecimento.

Desse modo, um conceito possui relação guiada pela razão entre outros conceitos, não de forma infinita, pois os fundamentos do conhecimento findam-se em um conceito, essas representações, expressas na forma de conceitos, possuem relação entre fundamentos da mesma classe, o que remonta a noção de domínio, logo, essa relação entre conceitos diz respeito à um determinado domínio do conhecimento, que é apresentado na forma de conceitos vinculados.

O que na Ciência da Informação poderia ser apresentado na forma de vocabulários controlados. Esta concepção de conceitos vinculados é bastante similar à adoção de termos significativos, que representam conceitos na elaboração de vocabulários para os domínios do conhecimento. Schopenhauer (2005) explica, que apesar dos conceitos possuírem essa

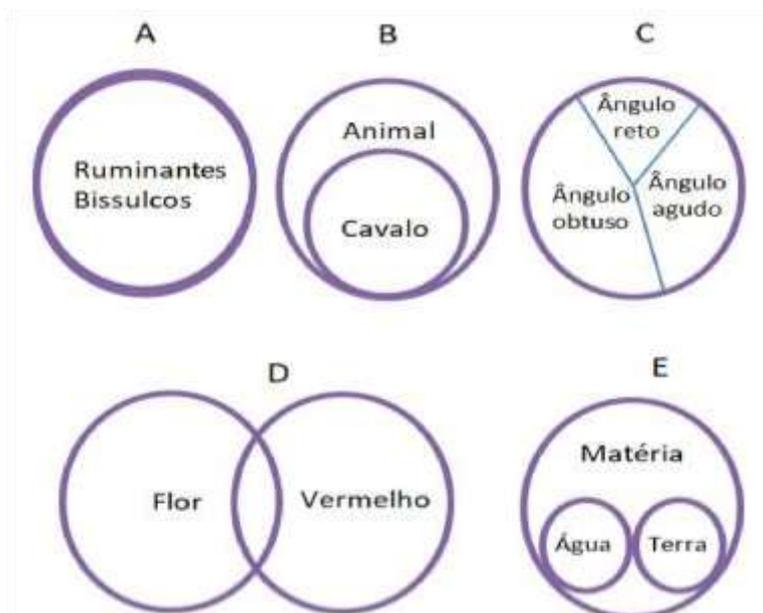
característica de representar um objeto, nem sempre ele será a representação de uma representação, isto é, possuir sua essência inteira e exclusivamente em referência à outra representação.

Existe a possibilidade de um conceito abranger mais de um objeto específico, devido a sua característica essencial, que é a generalização, “[...] não por que é abstraído de muitos objetos, mas ao contrário, justamente por que há generalidade, ou seja, a não determinação do particular, é essencial ao conceito como representação abstrata da razão, apenas por isso, diversas coisas podem ser pensadas mediante um mesmo conceito.” (SCHOPENHAUER, 2005, p. 89).

Esta característica universal dos conceitos possibilita esse tipo de ação, no entanto, um objeto não deve possuir dois conceitos distintos, pois um deles seria falso, portanto, necessitaria de algum elemento comum para vinculá-lo ao objeto. Porque Schopenhauer (2005) explica que a esfera de um conceito tem algo em comum com a esfera de outros conceitos, ou seja, que em parte, é nele pensado a mesma coisa que é pensado nos outros e vice-versa.

Um conceito vinculado a outro pode ser estruturalmente montado de acordo com os fundamentos da Organização e Representação do Conhecimento na construção de vocabulários controlados, por meio de termos, na sua forma de relacionar-se uns com os outros. Esta relação entre os conceitos segundo a concepção de Schopenhauer (2005), pode ser visualizada na figura 1, expressa da seguinte forma:

- a) Esferas de dois conceitos iguais, conceitos intercambiáveis expostos por um único círculo, a significar tanto um como outro;
- b) A esfera de um conceito encerra no todo a esfera de outro conceito;
- c) Uma esfera encerra duas ou mais esferas que se excluem e, ao mesmo tempo, preenchem a esfera;
- d) Duas esferas encerram, cada uma, parte da outra;
- e) Duas esferas estão contidas numa terceira; mas não a preenchem.

Figura 1 – Esferas de conceitos de Schopenhauer

Fonte: Adaptado de Schopenhauer (2005)

O arranjo estrutural das esferas de conceitos recupera um princípio matemático encontrado na teoria dos conjuntos, onde é possível unir, interseccionar, conter ou não, pertencer ou não, diferenciar, complementar, etc., determinados números. Esta noção matemática não é mera coincidência, pois Schopenhauer (2005) cita a contribuição Leonhard Euler (1707-1783), que levou esse procedimento a bom termo, usando círculos e, por isso, baseia-se nele.

Esta visão das esferas de conceitos evidencia uma relação de conjuntos e subconjuntos, união e intersecção de conceitos, que pode ser consultada nos Diagramas de Venn¹, por exemplo. Quando muitas esferas se arranjam dessa maneira, nasce uma longa cadeia de silogismos, esquemas de conceitos (SCHOPENHAUER, 2005). Entende-se que, essas esferas de conceitos, submetem-se à Lógica; portanto, o conceito não é apenas um exercício de abstração, mas um refinamento da razão.

Uma vez entregue à razão, outros fatores entram em cena no estudo do conceito, entre eles a linguagem e a Lógica; neste sentido, o que ora parecia abstrato, enquanto representação, ou seja, o conhecimento, começa a sofrer uma simbiose com realidade, isto é, por meio da linguagem. Não se trata apenas de uma sequência de palavras, mas da estruturação

¹ É uma forma de representar conjuntos e declarações que expressam relacionamentos entre eles, como união e intersecção.

coesa da abstração, função que poderia ser empregada à Lógica, interpretada como teoria do pensamento correto.

Neste ponto da discussão, é exposta uma intrínseca relação entre conceito, lógica e linguagem, esta reflexão reforça a necessidade de se perceber a extensão dos fundamentos empregados nos KOS, que comumente é reduzida ao conhecimento bibliográfico; portanto, há uma preocupação em evidenciar a gênese da representação e, conseqüentemente, do conceito.

Os conceitos são elementos-chave da representação, pois abarcam a relação sujeito e objeto e, ao mesmo tempo, expressam as abstrações contidas no consciente humano. Passível de aplicações lógicas, os conceitos se aplicam como substratos para o uso de uma linguagem, na qual é possível representar os objetos de todos os tipos, sejam eles materiais ou imateriais. A Lógica é uma disciplina que oferece suporte à formulação de conceitos, portanto é necessário apropriar-se de algumas noções que facilitem seu entendimento, assim como alguns princípios matemáticos que possuem valorosas explicações conceituais.

O uso da lógica para determinar um conceito, é um recurso utilizado por Dahlberg (1978) na construção de sua teoria. Para entender o conceito como unidade do conhecimento, apresenta a definição precípua do que deve ser estritamente verdadeiro. Para definir um conceito, além de utilizar as primazias da Lógica, é necessário identificá-lo pela ótica da linguagem, pois segundo Dahlberg (1978), o conhecimento fixa-se através dos elementos da linguagem.

Novos conhecimentos apareceram com novos elementos linguísticos e, também através destes, tornaram-se mais claros e distintos (DAHLBERG, 1978). A linguagem, a partir de uma perspectiva lógica, condiciona um ambiente favorável para a definição de um conceito, pois influencia uma tendência de verdade por meio de argumentos lógicos. “Com a ajuda das linguagens naturais é possível formular enunciados a respeito tanto dos conceitos individuais como dos conceitos gerais. É em base a tais enunciados que elaboramos os conceitos relativos aos diversos objetos.” (DAHLBERG, 1978, p. 102); logo, cada enunciado verdadeiro representa um elemento do conceito. Para se referir aos enunciados citados por Dahlberg (1978), neste trabalho foi adotado o termo “proposição”.

Por uma proposição vou entender o que é comum a um conjunto de sentenças declarativas sinônimas. Neste sentido de ‘proposição’, duas sentenças vão expressar a mesma proposição se elas tiverem o mesmo significado. Assim, aqui, mais uma vez, assim como com os enunciados, o problema da sinonímia vai ter de ser enfrentado. Uma outra explicação, popular desde o surgimento da semântica de mundos possíveis para as lógicas modais, identifica uma proposição com o conjunto de mundos possíveis nos quais ela é verdadeira, ou com uma função de mundos possíveis em valores de verdade. (HAACK, 2002, p.116).

Dahlberg (1978) confirma o problema da Filosofia encontrado por Wittgenstein (1968), que classificava como problema de linguagem; portanto, este seria sanado quando submetido às leis da Lógica. De acordo com Wittgenstein (1968), representar na linguagem algo que "contrarie as leis lógicas" é tão pouco possível como representar, na geometria, por meio de suas coordenadas, uma figura que contrarie as leis do espaço. "É perfeitamente possível representar um estado de coisas espacial contrário às leis da física, nunca, porém, contrário às leis da geometria. Um pensamento correto a priori seria aquele cuja possibilidade condiciona sua verdade." (WITTGENSTEIN, 1968, p.61).

Ao definir um conceito, é notável que este carregará um valor de verdade estritamente coeso. Todavia, Dahlberg (1978) beneficia-se de uma lógica aristotélica, que é carregada de silogismos, e que segundo Schopenhauer (2005), pode conduzir a erros, sim, produzem paralogismos e sofismas que se originam muito mais facilmente e são muito mais difíceis de resolver do que a ilusão de uma intuição sensível. Portanto, sempre resultam em uma conclusão verdadeira ou pelo menos válida; dessa forma, o uso da lógica em enunciados para definir conceitos, condiciona uma verdade. Esta concepção dá ao conceito a função de verdade que fundamenta um conhecimento, segue então como uma unidade básica e universal.

Segundo Dahlberg (1978), todo enunciado sobre objetos contém um elemento do respectivo conceito, quando relacionado à questão do aumento da produção de conhecimento, Dahlberg (1978) afirma que isto não implicará no alargamento do mesmo, mas na criação de novos conceitos. A questão torna-se paradigmática quando confrontada com o pensamento de Wittgenstein (1968, p.56) pois, "[...] se conheço o objeto, também conheço todas as possibilidades de seu aparecimento em estados de coisas. (Cada uma dessas possibilidades deve estar na natureza do objeto). Não é possível posteriormente encontrar nova possibilidade."

Essa visão paradigmática e até mesmo paradoxal, requer uma discussão cautelosa, pois esta dicotomia só se aplica em conceitos gerais, que necessitam de definições, portanto não necessitaria de um alargamento caso o conhecimento fosse modificado, mas implicaria no surgimento de um novo conceito. Wittgenstein (1968) é bem preciso quando relaciona o conhecimento de um objeto e suas possibilidades determinadas, apesar de não explicitar uma discussão sobre conceito. Dahlberg (1978) parte de uma visão aristotélica, na qual considera uma substância e seus acidentes. Ou seja, por meio da substância, é possível prever seus acidentes, pois, na esfera lógica, os acidentes estão vinculados à substância, assim como as possibilidades estão para um objeto.

Portanto, a relação entre objeto e conceito de Dahlberg (1978) é essencialmente afetada na lógica de Wittgenstein, pois, dentre outros fatores apontados por Dahlberg (1978), é fácil verificar que o conceito é constituído por elementos que se articulam numa unidade estruturada; por sua vez, o conhecimento seria um encadeamento desses conceitos. Na perspectiva de Wittgenstein (1968), o estado de coisas é uma ligação de objetos (coisas), pois é essencial para a coisa poder ser parte constituinte de um estado de coisas.

Nessa perspectiva, aponta que nada é acidental na lógica, pois, considerando o objeto ou a substância, é possível calcular tudo que estará vinculado a eles. O tempo e o espaço condicionam um objeto a unidades quantificáveis; portanto, o que está fora desses parâmetros não possui qualquer ligação com o objeto. Uma vez submetido às mesmas condições lógicas, não podemos pensá-lo fora da possibilidade dessa ligação (WITTGENSTEIN, 1968).

Só se houver objetos, pode haver forma fixa do mundo. O fixo, o subsistente e o objeto são um só. O objeto é o fixo, o subsistente; a configuração é o mutável, o instável. A configuração dos objetos forma o estado de coisas. No estado de coisas os objetos se ligam uns aos outros como elos de uma cadeia. No estado de coisas os objetos estão uns em relação aos outros de um modo determinado. (WITTGENSTEIN, 1968, p.58).

O que, em uma visão lógica, quer dizer que a definição de um conceito geral prever todos os outros conceitos associados a ele, pois estão relacionados a um objeto. Ou seja, em perspectiva lógica, um conceito geral determina todos os outros que se relacionam com ele, não havendo necessidade de incluir nenhum a mais. Isso reforça a visão pragmática na construção de um conhecimento aplicado e aponta a necessidade de algumas revisões teóricas. Pois, na visão de Wittgenstein (1968), na proposição está contido o signo, e o que no signo não vem expresso é indicado pela aplicação; o que os signos escondem, a aplicação exprime.

Para melhor explicar, Wittgenstein (1968) assume que, a partir das regras dos signos, construamos uma proposição lógica, e cada proposição da lógica é um *modus ponens* representado num signo, ou seja, sempre se pode conceber a lógica de tal modo que cada proposição seja sua própria prova. Wittgenstein usa o recurso do *modus ponens* assim como a Navalha de Occan para simplificar as inferências a que chegara. Pressuposto que elabora a partir da *Theory of types*, na qual ressalta que o erro de Bertrand Russell se revela quando, ao elaborar as regras dos signos, apela para a denotação desses signos.

Nenhuma proposição pode acertar algo sobre si mesma, pois o signo proposicional não pode estar contido em si mesmo (aí está toda a *Theory of types*). Uma função por isso não pode ser seu próprio argumento, pois o signo da função já contém a protofiguração de seu argumento, e não contém a si própria. (WITTGENSTEIN, 1968, p.68).

Ao considerar que no signo de uma proposição há traços essenciais (ligados a substâncias) e acidentais, que derivam da maneira particular de produzir o signo proposicional;

essenciais, aqueles que sozinhos tornam a proposição capaz de exprimir seu sentido (WITTGENSTEIN, 1968). Dahlberg (1978) destaca que, sempre que diferentes conceitos possuem características idênticas, deve-se admitir que entre eles existem relações. Neste sentido aponta as seguintes relações lógicas, hierárquicas (implicação), partitivas, de oposição (negação) e funcional (intersecção).

Apesar desse direcionamento que Dahlberg (1978) dá quando um conceito possui característica comum com outro, nota-se que ambos denotam sobre o mesmo objeto, sinonimamente ou em contextos diferentes. Wittgenstein (1968) aponta que na linguagem, corrente amiúde acontece, que a mesma palavra designa de modos diferentes, pertencendo a símbolos diferentes, ou ainda duas palavras, que designam de modos diferentes, são empregadas na mesma proposição.

Portanto, é neste sentido que Wittgenstein (1968) propõe o uso de uma linguagem simbólica que se submeta às regras da Lógica, para evitar esse tipo de confusão. A adoção de uma linguagem formal consideraria uma sintaxe lógica na construção das proposições, cujos signos determinariam uma forma lógica. Mesmo parecendo uma proposta instigante, no que diz respeito a eliminação de problemas desse tipo, tal adoção não consideraria a evolução da linguagem, considerando padrões pouco flexíveis a que seria submetida.

De acordo com Schopenhauer (2005), na linguagem e escritura, os signos dos conceitos possibilitam que uma precisa informação chegue a nós sobre cada coisa e cada relação, não importando quando e onde tenham ocorrido, porque poucos conceitos representam uma infinidade de coisas e estados. Os conceitos possuem esta característica generalista, na qual é possível completar uma informação orientada pelo sentido lógico que o conceito expressa; logo, a capacidade unívoca é essencial para o uso da linguagem na representação do conhecimento.

Os fundamentos sobre os quais estão erguidas as novas perspectivas de organização e representação do conhecimento, possuem intrínseca relação com o uso de novas tecnologias digitais. Sobretudo na Web Semântica, onde os vocabulários são constituídos por termos que representam conceitos, possíveis relacionamentos entre eles e as limitações no seu uso (CATARINO; SOUZA, 2012).

Os KOS são estruturas com preditivas lógicas que incorporam esses fundamentos, assim como o princípio elementar da representação, que é o sujeito e o objeto, encontrado como referência da tríplice relação no Framework de Descrição de Recursos para dados vinculados.

Além disso, a ligação conceitual é percebida por meio de termos significativos que abrangem determinada área do conhecimento. Quando aplicada ao SKOS, esta conexão abstrata preserva a conceptualização, porque possui um endereço lógico, que pode ser acessado por uma infraestrutura tecnológica de navegação.

Os conceitos como unidades do conhecimento representadas por termos significativos, possuem relações entre si, que podem ser simples ou complexas. Estes relacionamentos são configurados dentro de um sistema de organização, cujo objetivo é estabelecer determinadas categorias aos conceitos. Estes elementos tornam-se importantes fatores para se conhecer os processos de classificação, por exemplo, como a classificação de dois pontos (RANGANATHAN, 1939), que fundamenta-se no processo de categorização para se definir as facetas que determinarão um tipo de objeto.

Os conceitos estão em uma esfera abstrata, tal como as formas geométricas, Schopenhauer (2005) utiliza vários exemplos da matemática para explicitar a natureza dos conceitos. Por exemplo, um triângulo, losango ou dodecaedro é resultado de um cálculo que indica o formato de um objeto no espaço, até então não há formas de se provar a origem natural dessas formas senão pela própria Geometria, porque elas são resultado do refinamento da razão.

As representações que fundamentam o entendimento humano sobre o funcionamento da realidade, são formas desenvolvidas para codificar a natureza e as relações entre os seres para uma linguagem assimilável à razão. Os conceitos são os principais elementos do processo de representação, pois abarcam a relação sujeito e objeto e ao mesmo tempo expressam as abstrações contidas no consciente humano. Orientados por aplicações lógicas, os conceitos se aplicam como substratos para o uso de uma linguagem, na qual é possível representar os objetos de todos os tipos, sejam eles materiais ou imateriais.

O conhecimento científico carrega o fardo da verdade, daquilo que pode ser provado e explicado por meio de um método, portanto os conceitos científicos carregam uma boa dose de métodos de tendências cartesianas e positivistas. Esta é uma das características da ciência, pois seu conhecimento deve ser amplamente discutido, analisado, refutado ou corroborado, pois é isso que tem caracterizado seu progresso desde então, assim como as diferentes correntes teóricas que originaram os tipos de conhecimento.

O racionalismo é um dos grandes produtores de conhecimento, pois satisfaz a conjectura das ideias e a adoção de métodos filosóficos, como dedutivo e indutivo, provocando uma fidedigna sensação de identificação da verdade. A Lógica, grande aliada desse movimento,

retoma a adoção de silogismos que potencializam qualquer distinção entre certo e errado, apesar de não oferecer, substancialmente, a verdade absoluta das coisas. Nesta perspectiva, outros tipos de asserções sobre a produção de um conhecimento verdadeiro são desenhados na história da ciência, que alcança os pilares da modernidade sobre várias nuances conceituais.

As respostas apresentadas pela ciência são decorrentes de análises empíricas e racionais, que advêm de um constructo conceitual que se entende pelo método científico, que se configura como o caminho percorrido pelo cientista para a obtenção de respostas. O método é o grande alicerce do cientista, por meio dele é possível verificar tais procedimentos e o grau de confiabilidade dos resultados devidamente validados pelos pares. Por meio deles é possível estabelecer leis, hipóteses e conceitos a respeito dos fenômenos; a adoção de um método implica a operacionalização mental das ideias por meio de várias etapas lógicas que permitiram chegar a determinados conceitos.

René Descartes é um grande contribuinte às ciências ao apresentar, em sua obra “*Discurso do método*”, uma sugestão para estruturar as ideias que extraíam possíveis respostas que o direcionem a um conhecimento verdadeiro. Descartes não estabelece um modelo formal, mas apresenta fases que se deve passar para esclarecer o julgo dos argumentos que favorecem a percepção errônea da verdade, nisto presume que afastará preceitos e o senso comum que influenciam na estruturação dessas ideias. Na segunda parte do Discurso, Descartes esboça uma nota crítica sobre a Lógica que parece muito oportuna.

[...] quanto à lógica, seus silogismo e a maior parte de suas outras instruções servem mais para explicar aos outros as coisas que se sabem, ou mesmo como a arte de Lúlio, para falar sem discernimento daquelas que se ignoram, do que para aprendê-las; e, embora contenha efetivamente preceitos muito verdadeiros e muito bons, existem, misturados a eles, tantos outros que são nocivos ou supérfluos, que é quase tão difícil separá-los quanto tirar uma Diana ou uma Minerva de um bloco de mármore ainda não esboçado. (DESCARTES, 2001, p.22).

Nessa crítica, Descartes aponta algumas limitações da Lógica para obter repostas significativas, pois, embora se abasteça de premissas verdadeiras, estas podem induzir a um pensamento correto, mas de difícil apreensão. Por exemplo, é mais fácil imaginar um objeto com três dimensões do que com onze. A ordenação das ideias é ponto de partida que vai do simples para o mais complexo, indica uma forma elementar adotada por Descartes (2001), preceitos que ele afirma serem empregados pelos geômetras para se chegar às mais difíceis demonstrações.

Descartes (2001) acreditava que a adoção do método criado por ele o conduziria ao mais próximo possível da verdade, desde que seguisse cabalmente cada etapa, além de manter

suas máximas. Com isso, revela uma jornada solitária guiada apenas pela razão e o método que lhe permitiu caminhar, com intuito de progredir em tantas ciências que não se satisfizessem com as verdades que não fossem minuciosamente colocadas à prova da razão.

Augusto Comte (1791-1857), precursor do positivismo, contemporâneo do Iluminismo, tem, como pressuposto basilar, a razão; e, a Ciência, como meio de chegar à verdade sobre a vida, os fatos e os fenômenos. De acordo com Sell (2001), o positivismo, em sua dimensão filosófica, corresponde a ideia de que a ciência é a única explicação legítima para a realidade. É nesta perspectiva que Comte chega à Lei dos três estágios, a saber: estado teológico, metafísico (filosófico) e o positivo (científico).

O estado teológico é fase primeira das explicações a respeito dos fenômenos, comumente atribuídos a personalidades divinas; uma espécie de conhecimento mítico aplicado aos eventos naturais do curso humano. O conhecimento teológico, a partir da religião, é a fonte produtora das explicações a respeito da vida e dos fenômenos naturais (SELL, 2001). Uma fonte que passa a ser questionada após a cisão da ciência com a religião, que começa a questionar tais fundamentações que, do ponto de vista racional e lógico, passam a conter falhas pela adoção de dogmas como argumentos incontestáveis.

No estado metafísico, o conhecimento, originado para explicações dos fenômenos, é decorrente de profunda abstração sobre a causa e os efeitos que lhe são atribuídos. A explicação transcenderia a materialidade essencialmente contida no objeto, esta relação não é mensurável materialmente, mas seria completamente explicada no mundo das ideias, que disporia de condições necessárias para o entendimento nuclear do fenômeno. Sell (2001) destaca que as causas divinas são substituídas por causas mais gerais – as entidades metafísicas – buscando, nestas entidades abstratas (ideias), explicações sobre a natureza das coisas e a causa dos acontecimentos.

O estado positivo (científico) seria a junção do empirismo e do racionalismo como fonte de extração do conhecimento para explicações dos fenômenos e, por meio de métodos indutivos e dedutivos, desenvolveria uma hipótese que alicerçaria tal conhecimento ao ponto de se considerar a verdade sobre os objetos estudados. Para Sell (2001), no estado positivo ou científico, as causas primeiras e absolutas são substituídas pela observação da relação entre os fenômenos, mediante a rigorosa pesquisa científica.

Augusto Comte apresenta um modelo para explicações dos fenômenos, sejam eles naturais ou sociais, importado das ciências natureza, como a Matemática, a Física e a Química;

que, por meio de um método científico, obteria uma explicação primária, pois seria possível identificar padrões lógicos e quantificáveis, que se submeteriam a um método. Este processo de extração da verdade condiciona a unicidade da explicação causal dos fenômenos, por meio de provas comprobatórias e plenamente racionais.

Precisamente foi o positivismo lógico que formulou o célebre princípio da verificação (demonstração da verdade). Segundo este princípio, será verdadeiro aquilo que é empiricamente verificável, isto é, toda afirmação sobre o mundo deve ser confrontada com o dado. (TRIVIÑOS, 1987, p.37).

Triviños (1987) explica que o positivismo não aceita outra realidade que não seja a dos fatos, fatos que possam ser observados. A partir disso, é possível analisar, por meio de métodos que consigam sustentar determinadas hipóteses. Triviños (1987) ressalta que facilmente se observa que a filosofia positivista se colocou no extremo oposto da especulação pura, exaltando, sobretudo, os fatos. Nesta visão, o mundo das ideias explorado na metafísica de Kant, aparece como um estado intermediário, que proporciona o estado positivo, que seria suficiente para explicação dos fenômenos que compunham a realidade dos fatos.

A filosofia positivista destaca o percurso normal da Ciência para consolidação do conhecimento científico e suas teorias. Todavia, Kuhn (1998) ressalta que muitas descobertas causaram ou contribuíram para a mudança de paradigmas, e isto só foi possível quando algumas crenças, ou procedimentos anteriormente aceitos, foram descartados e, simultaneamente, substituídos por outros. Isso quer dizer que, embora a Ciência ofereça uma certa estabilidade sobre o conhecimento gerado a partir de métodos científicos, é possível que esse conhecimento seja reconstruído a partir de novos métodos.

Este cenário de mudança científica e sua tendência de construção e desconstrução, favoreceu as críticas sobre a cristalização do conhecimento e seus conceitos. Deleuze e Guattari (1996, p.52) explicam que, “os conceitos são superfícies ou volumes absolutos, disformes e fragmentários, enquanto o plano é o absoluto ilimitado, informe, nem superfície nem volume, mas sempre fractal”. Essa fragmentação é o combustível para as transformações conceituais, ou seja, embora os conceitos transmitam essa noção de estabilidade, a mudança de paradigmas pode desencadear novos panoramas do conhecimento.

A discussão filosófica sobre o conhecimento e sua origem, como pode ser vista, possui uma intrínseca relação com a Teoria da Ciência, pois esta possui uma ligação direta com a Teoria do Conhecimento, esta, por sua vez, ocupa-se do conhecimento humano sobre as coisas (HESSEN, 2003). Os estudos no âmbito da Organização do Conhecimento não só pode, como deve, realizar imersões teóricas sobre a origem dos conceitos e do conhecimento, para que

aproximem cada vez mais da metáfora citada por Sócrates, a respeito da obra de Heráclito quando se refere aos mergulhadores de Delos², pois o estudo dos conceitos não se limitam a sua instrumentalização em KOS, mas a reflexões epistemológicas que questionam sua própria natureza.

As reflexões filosóficas sobre as teorias que alimentam a Organização do Conhecimento precisam ser revisitadas constantemente, para que seja explicitado o motivo de sua existência e não sirva apenas como hábito científico que se cristalizou em discurso. O conhecimento e a informação são objetos de estudos imateriais que podem ser instrumentalizados, enquanto os conceitos são partes da realidade como a conhecemos, neles estão as definições mais abrangentes sobre representação das coisas no mundo. Os conceitos não estão nem lá nem cá, são formas abstratas que podem estar em todos os lugares e em lugar nenhum, o universo da abstração é o que os sustém.

O conhecimento é o resultado da combinação semântica de conceitos, e os Sistemas de Organização do Conhecimento são instrumentos que viabilizam sua gestão. De acordo com Hjørland (2015), os conceitos são carregados de teoria, às vezes é difícil identificar as teorias, que estão envolvidas na especificação do significado de um determinado termo, mas elas estão lá. Por isso é importante o estudo das teorias, especialmente aquelas que orientam instrumentos de representação como os KOS.

Se para Dahlberg (2006), organização do conhecimento é a ciência que ordena a estruturação e sistematização dos conceitos, entender desde a composição gênica, em nível epistêmico, até esquemas de classificação, vocabulários controlados, entre outros incluídos nos KOS com uso de novas tecnologias, é um desafio que precisa ser enfrentado pela Ciência da Informação, para dar organicidade teórica aos instrumentos aplicados nesse processo.

2.2 Knowledge Organization Systems

As primeiras formas de organização foram expressas na forma de esquemas de classificação geral, porque reuniam o conhecimento de diferentes áreas para gerar uma ordenação lógica que atendesse a esfera conceitual e aplicada. De acordo com Barbosa (1969), as classificações bibliográficas tiveram raízes na classificação dos conhecimentos humanos, tendo sido Platão o primeiro a agrupá-los segundo bases filosóficas. Para Ferreira, Maculan e Naves (2017), a teoria da classificação tem aplicações em todos os ramos do conhecimento.

² Exímios mergulhadores de águas profundas da ilha de Delos na Grécia, também conhecidos como caçadores de tesouros.

Esses fundamentos teóricos estão diretamente relacionados aos diversos tipos de sistemas e aplicações desenvolvidos ao longo do tempo; por exemplo, a árvore de Porfírio, que apresenta um tipo de classificação binária, na qual a divisão do conhecimento parte, de termos de grande extensão e pouca intensão, para termos de pouca extensão e grande intensão. Este modelo de classificação do ano 305 assemelha-se a um método de classificação computacional que utiliza uma estrutura denominada árvore de decisão³.

Este entrelaçamento é perfeitamente comum nos KOS, pois sua natureza é teórica/aplicada, e, segundo Hodge (2000), engloba dicionários, glossários, arquivos de autoridade, taxonomias, sistemas de classificação, cabeçalhos de assunto, tesouros, redes semânticas e ontologias. Esses instrumentos, utilizados para organizar o conhecimento, reúnem características que possibilitam, sobretudo, o estabelecimento de um processo de categorização dos conceitos.

Os KOS são resultados deste processo de reflexão, como dispositivos sistematicamente ordenados com preditivas lógicas e semânticas, que permitem o processo de representação com o uso de uma determinada linguagem. Desse modo, organização do conhecimento apresenta, em sua gênese, uma protofiguração com prospecção tecnológica. Logo, os instrumentos da organização do conhecimento são os principais mediadores dessa transformação canônica figurativa, revelando os avanços que a tecnologia subsidia nos diversos campos de estudo.

A organização do conhecimento representa uma das principais contribuições intelectuais que estabelece o campo de estudo dos KOS (HJØRLAND, 2016). A fundamentação teórica-filosófica da organização do conhecimento possibilita uma visão holística dos instrumentos de representação. Pois permite a manipulação desses instrumentos em novas demandas que possam surgir, diferentes do contexto de criação, mas igualmente necessárias, cujas respostas podem estar contidas em determinados KOS.

A linguagem é o grande elo que vincula o processo de categorização dos conceitos, a representação do conhecimento e os KOS propriamente ditos, por meio dela é possível criar formas de manuseá-los para atender as demandas do processo de representação. O controle terminológico é um dos mecanismos encontrados para lidar com este aspecto do conhecimento. Nesse sentido, os instrumentos de controle terminológico fundamentam-se em estruturas lógicas e semânticas que norteiam a organização e representação do conhecimento.

³ Modelos estatísticos utilizados para treinamento supervisionado de classificação e previsão de dados.

As relações significativas da linguagem são resultados das formas propositivas dos conceitos expressos na forma de enunciados, o uso de premissas e conectivos lógicos são elementos participativos na construção de proposições falsas, verdadeiras, axiomas, paradoxos, entre outros, neste aspecto, o campo da Organização do conhecimento apreende os elementos e estruturas dos sistemas conceituais (OHLY, 2018).

A linguagem é uma ampla esfera que interage com diversos campos disciplinares e que subsidia as discussões de esferas interdisciplinares, como é o caso da Ciência da Informação. Neste sentido, estudos correlacionados, de ambas as áreas, permitem intersecções que podem aproximar ou gerar campos independentes, como a Terminologia, que concebe sua teoria geral por meio da tese de doutorado de Eugen Wüster, em 1931 (WÜSTER, 1998).

Segundo Almeida (2003), Wüster acabou dando à Teoria Geral da Terminologia (TGT), a configuração de uma terminologia representativa, já que era preponderante denominar e etiquetar a informação, pois as terminologias precisavam ser controladas para que a comunicação fosse inequívoca e eficaz.

O controle terminológico para a representação da informação e do conhecimento, envolve não apenas questões de ordem morfossintática, porém de estruturação lógica. O trabalho de Wüster influenciou os padrões de termos, hoje elencados pela *International Standardization Organization* (ISO), que atualmente desenvolve normas técnicas que cobrem várias áreas do espectro humano, como informação, meio ambiente, qualidade nos processos, entre outros.

No Brasil, a representação e tradução das normas da ISO são realizadas pela Associação Brasileira de Normas Técnicas (ABNT), por meio das Normas Brasileiras de Regulamentação (NBRs). Essas normas têm como objetivo padronizar determinados processos, no entanto, o grande desafio é proposto quando se trata da linguagem humana, pois é notável sua transmutação e seu multilinguismo; a Terminologia, enquanto ciência independente e fundamentada somente na TGT, possui suas fragilidades, apontadas por especialistas como Teresa Cabré, que desenvolve a Teoria Comunicativa da Terminologia como alternativa de equilíbrio teórico fundamental.

A Terminologia é concebida como um campo interdisciplinar construído a partir de três teorias, que, da mais remota a mais próxima seriam as seguintes: uma teoria do conhecimento que explique como se conceitua a realidade[...]; uma teoria da comunicação que descreva a partir de critérios explícitos dos tipos de situações que podem produzir [...]; uma teoria da linguagem que dê conta das unidades

terminológicas propriamente ditas dentro da linguagem natural [...]. (CABRÉ, 1999, p.100-101, tradução nossa).

A Terminologia é uma disciplina que tem como objeto de estudo a estruturação de termos especializados em determinados domínios do conhecimento, cuja ascendência remonta a elaboração de vocabulários controlados, cabeçalhos de assuntos, glossários e tesouros. Isso não implica dizer que a linguagem e o conhecimento estão sobrepostos em balanças para medir o grau de importância de cada um, mas que há uma relação de dependência entre essas duas esferas, onde o que importa não é a predominância de uma ou outra, mas a inserção de ferramentas e processos que consigam integrá-los; a fim de estabelecer parâmetros para o controle terminológico em determinados domínios do conhecimento.

Os KOS possuem instigantes aplicações nas tecnologias digitais, pois envolvem soluções muito práticas para representação da informação, principalmente no ambiente web, com o uso de padrões interoperáveis. A indexação é um dos princípios básicos utilizados nesta área, por meio do qual é possível recuperar vários tipos de arquivos em uma busca por palavras-chave no ambiente web.

Nesse aspecto, interoperabilidade entre sistemas de representação e recuperação da informação, favorece a efetivação de um recurso oferecido. Segundo a NISO (2004), interoperabilidade é a capacidade de vários sistemas, com diferentes plataformas de hardware e software, estruturas de dados e interfaces, para trocar dados com perda mínima de conteúdo e funcionalidade. No âmbito gerencial, de acordo com a ENAP:

A interoperabilidade pode ser entendida como uma característica que se refere à capacidade de diversos sistemas e organizações trabalharem em conjunto (interoperar) de modo a garantir que pessoas, organizações e sistemas computacionais interajam para trocar informações de maneira eficaz e eficiente. (BRASIL, 2012, p.06).

Neste sentido, a noção de interoperabilidade semântica é entendida como a capacidade de sistemas distintos executarem operações de troca de informações entre si. Para a ENAP, a interoperabilidade semântica garante que os dados trocados tenham seu significado corretamente interpretado dentro do contexto de uma dada transação ou busca de informação. (BRASIL, 2012). Um exemplo aplicável dos conceitos de interoperabilidade semântica, citado pela ENAP, é o Vocabulário Controlado do Governo Eletrônico (VCGE), disponível nos formatos JSON, SKOS (RDF/N3, RDF/XML, RDF/NT).

O tesouro é um exemplo de aplicação semântica, e pode ser definido como um vocabulário de termos relacionados genérica e semanticamente sobre determinada área de conhecimento (MOTTA, 1987), com termos genéricos, específicos e relacionados entre si. Os

processos de relacionamento entre termos que ora encontravam-se disponíveis apenas em materiais impressos, pelo uso das tecnologias digitais vigoram atualmente sob a ótica de vocabulários vinculados na web.

Os termos, que outrora folheados em tesouros impressos, podem ser acessados com um clique, por meio de modelos de organização do conhecimento que se fundamentam nessa mesma estrutura de representação da informação aplicada ao ambiente web. Dentre as características interoperáveis dos vocabulários, destaca-se o uso de modelos de descrição padronizados, que podem ser adequadamente representados por formatos que consigam agrupar termos, hierarquias e relações entre si, um exemplo disso é a *Basel Register of Thesauri, Ontologies & Classifications* (BARTOC), um banco de dados com registros de KOS desenvolvido pela Biblioteca da Universidade de Basel, na Suíça.

Controlar a linguagem não é uma tarefa simples, por isso, os vocabulários controlados em sua maioria são de áreas ou assuntos específicos. “[...] Um vocabulário controlado é essencialmente uma lista de termos autorizados para controlar sinônimos, diferenciar homógrafos e estabelecer as relações hierárquicas e não hierárquicas”. (LANCASTER, 2004, p. 19). Os vocabulários controlados destinam-se à padronização de termos representativos, além de estabelecer relações entre esses termos e poder ser classificado pelo seu nível de complexidade, como cabeçalhos de assunto, glossários, taxonomias, tesouros, etc. Cada vocabulário possui características que podem se adaptar a diferentes necessidades. O *Networked Knowledge Organization System and Service* (NKOS), Guarino (1998) e Mazzocchi (2018), oferecem definições que podem acolher o entendimento desses vocabulários contidos nos KOS.

Quadro 1 – KOS- Vocabulários e suas definições

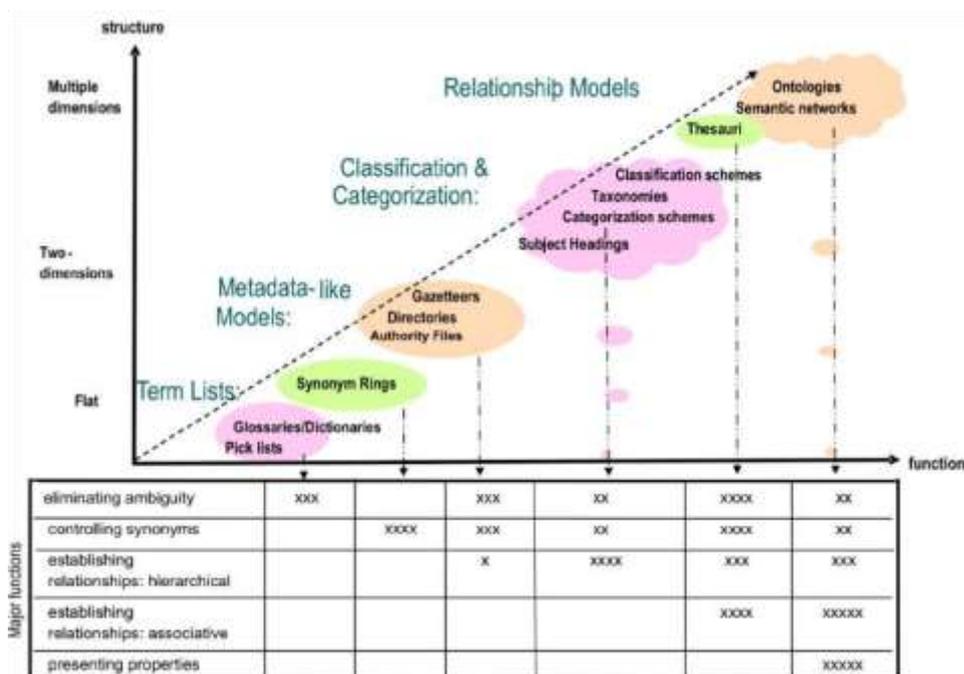
Vocabulário Controlado	Definição
Anéis de sinônimo	Conjuntos de termos considerados equivalentes para fins de recuperação de informações (IR).
Glossários	Um glossário é uma lista de termos, geralmente com definições. Os termos podem ser de um campo de assunto específico ou aqueles usados em um determinado trabalho. Os termos são definidos dentro deste ambiente específico e raramente têm significados variantes fornecidos.
Cabeçalhos de Assunto	Este esquema fornece um conjunto de termos controlados para representar os assuntos dos itens em uma coleção. As listas de cabeçalhos de assunto podem ser extensas, abrangendo uma ampla gama de assuntos. No entanto, a estrutura da lista de cabeçalhos de assunto é geralmente muito superficial, com uma estrutura hierárquica limitada. Em uso, os cabeçalhos de assuntos tendem a ser pré-coordenados, com regras sobre como os cabeçalhos de assuntos podem ser unidos para fornecer conceitos mais específicos.
Taxonomias	Refere-se a um esquema que apresenta biota em um arranjo hierárquico baseado em alguma característica. As taxonomias estão sendo cada vez mais usadas em projetos orientados a objetos e em sistemas de gerenciamento de conhecimento para indicar qualquer agrupamento de objetos com base em uma característica particular.
Tesouro	Esses KOS são baseados em conceitos e mostram relações entre os termos. Os relacionamentos, comumente expressos em um tesouro, incluem hierarquia, equivalência e associativo (ou relacionado). Essas relações são geralmente representadas pela notação BT (termo mais amplo), NT (termo mais restrito), SY (sinônimo) e RT (associativo ou relacionamentos associativos que podem ser mais granulares em alguns esquemas).
Ontologia	No sentido filosófico, uma "ontologia" é como um sistema particular de categorias que versa sobre uma certa visão do mundo. Dessa forma, esse sistema não depende de uma linguagem particular: a ontologia de Aristóteles é sempre a mesma, independente da linguagem usada para descrevê-la. Por outro lado, em seu uso mais prevalente em IA [inteligência artificial], uma ontologia é referida como um artefato de engenharia, constituído por um vocabulário específico usado para descrever uma certa realidade e um conjunto de pressupostos explícitos, relacionados com o significado pretendido para as palavras do vocabulário.
Redes semânticas	Sistemas onde os termos ou conceitos são modelados como em uma rede de tipos de relacionamentos variáveis; eles são mais ricos que os dicionários, na definição de categorias ou tipos semânticos e relações semânticas.

Fonte: elaborado a partir de NKOS (2000); Guarino (1998) e Mazzocchi (2018)

Os KOS, expressos na forma de vocabulários controlados, podem representar tanto estruturas simples (glossários ou dicionários), como sistemas complexos (redes semânticas e ontologias). A figura 2 representa os variados tipos de KOS, que de acordo com Zeng (2008), foram originados por Hodge (2000) e adotados pelo NKOS, conseqüentemente também adotados pela ISKO. Apresentam uma organização de acordo com o grau de controle introduzido (da linguagem natural à linguagem controlada) e a força de sua estrutura semântica

(de fracamente estruturada à fortemente estruturada), correspondendo às principais funções de KOS. (NKOS, 2000).

Figura 2 - Variedade de KOS



Fonte: Zeng (2008)

Zeng (2008) apresenta uma variedade de KOS explorada por dois eixos, que dizem respeito à estrutura e função de cada vocabulário. Quanto a estrutura, pode ser plana, com duas ou múltiplas dimensões, progressivamente apresentada por meio de lista de termos (glossários, dicionários); modelos de metadados (*gazetteers*, diretórios, arquivos autoritários); classificação e categorização (cabecinhos de assunto, esquemas de categorização, taxonomias, esquemas de classificação) e modelos de relacionamento (tesauros, redes semânticas, ontologias).

No que diz respeito à função, as listas de termos propõem a eliminação de ambiguidade, enquanto os modelos de metadados também se preocupam em eliminar ambiguidade, controlar sinônimos e estabelecer relacionamentos do tipo hierárquico. A categorização e classificação também englobam essas funções, já os modelos de relacionamento, além dessas, estabelecem relacionamentos do tipo associativo e apresentam suas propriedades.

Cada tipo de vocabulário cumpre uma função diferente, que pode limitar-se apenas a eliminação de ambiguidade, no caso dos glossários; ou cumprir essa função, acrescida do controle de sinônimos e/ou estabelecimento de relações hierárquicas e associativas, no caso dos

tesauros. De acordo com a National Information Standards Organization (2010), o processo de organizar uma lista de termos objetiva:

- a) Indicar qual dos dois ou mais termos são sinônimos autorizados para uso;
- b) Distinguir homógrafos;
- c) Indicar relações hierárquicas e associativas entre os termos no contexto de um vocabulário controlado.

Os vocabulários controlados são resultados da aplicação de regras semânticas e terminológicas, relacionados a um conjunto de termos selecionados em determinado domínio do conhecimento. As normas que permitem a elaboração dessas ferramentas, passam por constantes atualizações decorrentes dos avanços tecnológicos. Deste modo, os vocabulários passaram a oferecer modelos que permitem integrar um padrão eficiente para a recuperação da informação no ambiente web. Em virtude disso, os vocabulários controlados, segundo a National Information Standards Organization (2010), servem a cinco propósitos:

- 1. Tradução:** fornecer um meio para converter a linguagem natural de autores, indexadores e usuários, em um vocabulário que pode ser usado para indexação e recuperação;
- 2. Consistência:** promover a uniformidade no formato do termo e na atribuição de termos;
- 3. Indicação de relacionamentos:** indicar relações semânticas entre os termos;
- 4. Etiquetar e navegar:** fornecer hierarquias consistentes e claras em um sistema de navegação, para ajudar os usuários com a localização de objetos do conteúdo desejado;
- 5. Recuperação:** servir como auxílio de pesquisa na localização de objetos de conteúdo.

Para que estes propósitos sejam alcançados, a National Information Standards Organization (2010) fornece um conjunto de abreviações e acrônimos para indicar a posição e o relacionamento estabelecido entre os termos, apresentados no quadro 2. Segundo Clarke e Zeng (2012), embora a ISO tenha um objetivo claro de organizar conceitos e suas relações, essa a nomenclatura não foi inventada pela ISO 2788, nem pela norma nacional americana ANSI Z39.19 de 1974.

De acordo com Clarke e Zeng (2012), essas terminologias foram usadas em tesauros em toda a década de 1960, por conta de um trabalho intitulado Tesauro de Engenharia

e Termos Científicos (*Thesaurus of Engineering and Scientific Terms* – TEST), que adota a terminologia *Broader Term* (BT) para (Termo Genérico – TG), *Narrower Term* (NT) para (Termo Específico - TE) e *Related Term* (RT) para (Termo Relacionado – TR). O TEST foi o principal responsável por essa adoção terminológica em vocabulários controlados.

Quadro 2 – Terminologia para Vocabulários Controlados

Abreviação/ Sigla	Definição/Tradução
BT	Termo mais amplo
BTG	Termo mais amplo (genérico)
BTI	Termo mais amplo (instância)
BTP	Termo mais amplo (partitivo)
HN	Nota de História
NT	Termo mais restrito
NTG	Termo mais curto (genérico)
NTI	Termo mais curto (instância)
NTP	Termo mais curto (partitivo)
RT	Termo relacionado
SN	Nota de Escopo
TT	Termo Superior
U	USAR
UF	USADO PARA
X SN	Veja Nota de Escopo para.

Fonte: National Information Standards Organization (2010)

O processo de elaboração de um vocabulário controlado é pautado em várias etapas, que podem demandar muito tempo, pois o simples fato de tentar controlar algo tão suscetível a transformações, como a linguagem, exige um constante acompanhamento das mudanças implementadas, que pode ser desde a inserção de um novo termo, como a inclusão de novas tecnologias para impulsionar a eficiência de um vocabulário.

Tecnologias como os vocabulários vinculados têm se beneficiado desta fundamentação teórica em seu processo de publicação, pois permitem a conexão semântica entre termos que podem representar conceitos por meio de uma série de relacionamentos, facilitando a convergência interoperável desses vocabulários. Sobre este aspecto, os KOS oferecem os subsídios necessários para o entendimento das tecnologias semânticas que incorporam os vocabulários vinculados, que em síntese expressam o relacionamento contextual entre termos.

A tecnologia dialoga com os clássicos, pois o progresso fundamenta-se no percurso anterior como parte contribuinte do avanço. Na sociedade atual, o grande volume de dados,

informação e conhecimento, indica um fluxo intermitente de transformações em nível global. Essas transformações precisam, obviamente, fundamentar-se nas correntes teóricas que embasam seu surgimento.

O controle terminológico é um dos principais objetivos do processo de tradução da linguagem natural para a linguagem artificial/documentária, o vocabulário controlado possui importante grau de participação nesse processo, por meio dele são elencados quais termos representarão determinados conceitos. Uma vez selecionados, esses termos comporão um conjunto de descritores que representarão, tematicamente, determinado domínio do conhecimento.

Diversas iniciativas têm surgido, ao longo do tempo, para publicar vocabulários controlados, mas comumente expressos na forma de tesouros. Com o auxílio de novas tecnologias semânticas, projetos como o *The Food and Agriculture Organization Thesaurus* (AGROVOC) e *UNESKOS Vocabulary* são alguns casos que popularizaram os vocabulários vinculados na forma de tesouros.

A literatura científica também reforça o interesse em tesouros como importante ferramenta de aplicação de KOS no ambiente digital, conforme destaca Gray et al (2010); Schandl e Blumauer (2010); Pastor-Sánchez; Martínez-Méndez; & Rodríguez-Muñoz, (2012); Zapilko et al (2012); Catarino (2014); Ramalho (2015b); Suominen et al (2015); Jonquet et al (2018); Albertoni et al (2018); Frey; Accomazzi (2018). Neste sentido, o tesouro tem se mostrado como importante tipo de vocabulário que consegue intercambiar os fundamentos basilares dos KOS e aplicá-los no ambiente digital.

Um tesouro é um vocabulário controlado organizado em uma ordem conhecida e estruturado de modo que as diversas relações entre os termos sejam claramente identificadas por indicadores de relacionamento padronizados. Os indicadores de relacionamento devem ser utilizados reciprocamente. (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010, p.18, tradução nossa)

Para Cavalcanti (1978), uma lista estruturada de termos, associada e empregada por analistas de informação e indexadores, para descrever um documento com a desejada especificidade, em nível de entrada e para permitir aos pesquisadores a recuperação da informação que procura. Os tesouros representam um conjunto de conceitos expressos por um número controlado de descritores, o que torna bastante útil sua aplicação em áreas especializadas no processo de recuperação da informação. Eles têm a função de auxiliar na tradução da linguagem natural, utilizada pelo autor do texto e pelo usuário final, convertendo-

a em uma linguagem controlada, possibilitando a descrição padronizada de assuntos abordados em documentos (CAFÉ; BRASCHER; SUJII, 1990).

O uso das abreviaturas é uma ação comungada por um processo similar nos sistemas de classificação, que utilizam notações numéricas e sinais que objetivam a representação temática de um assunto, resultante da aplicação lógica de regras pré-estabelecidas entre os assuntos contidos nesses sistemas. Um tesouro deve apresentar e distinguir claramente as relações básicas que unem seus termos (AUSTIN, 1993). Ou seja, precisa ser um instrumento objetivo e lógico, respeitando as limitações terminológicas do domínio escolhido. “A simples adoção de um tesouro como uma linguagem artificial, controlada, já contribui em muito para a diminuição da inconsistência na recuperação da informação em uma base de dados.” (SOUTO, 2003, p. 78).

A seleção de candidatos a termos, deve abranger todos os contextos de uso dos descritores aplicados ao domínio selecionado, função primordial empregada será o relacionamento conceitual dos termos. Portanto, as regras de aplicações lógicas precisam ser bem definidas para que permitam a efetividade do processo de representação e recuperação da informação.

Ao longo dos últimos anos, inúmeras iniciativas têm sido propostas, com o intuito de desenvolver e aperfeiçoar métodos e ferramentas que possam potencializar o uso de linguagens e vocabulários controlados em ambientes computacionais, buscando favorecer um melhor atendimento às demandas informacionais contemporâneas, originando um crescente número de pesquisas relacionadas ao desenvolvimento e uso de KOS.

O uso de um determinado KOS para sistematizar o conhecimento produzido por um indivíduo ou instituição é o primeiro passo adotado neste processo, no entanto, logo nesta etapa podem surgir diversos entraves, como a compatibilidade desse sistema com outros, o que pode comprometer o seu funcionamento. Por isso, algumas soluções têm surgido para buscar a elementaridade dos processos, buscando elementos comuns e simples que podem ser utilizados como ponto de partida até alcançar estruturas mais complexas.

Com o auxílio de instrumentos tecnológicos isso tornou-se possível, pois o *Simple Knowledge Organization System* (SKOS) revigora fundamentos basilares da organização do conhecimento para representar, de forma dinâmica e precisa, os relacionamentos conceituais entre os termos expressos na forma de vocabulários, para que isso se tornasse possível foi necessário considerar a relação entre sujeito, predicado e objeto.

Apesar do predicado surgir como algo novo nesse processo, a relação entre sujeito e objeto é algo percebido desde Schopenhauer (2005), logo, a representação do ponto de vista tecnológico é apenas um diálogo com o tempo presente mediado pelas transformações da própria linguagem, portanto, estão intrinsecamente relacionados. Modelos de representação, como os *Functional Requirements for Bibliographic Records* (FRBR) e o BIBFRAME, são exemplos de como uma estrutura conceitual pode ser aplicada em instrumento de representação.

Não obstante, o SKOS surge como resultado de uma lacuna que põe teoria e prática em campos distintos, logo, incorpora fundamentos da Ciência da Informação para criar estruturas semânticas que estabelecem novos contextos no ambiente web. Neste sentido, sua exploração é uma atividade necessária, portanto, sua estrutura técnica e funcional precisa ser entendida de igual forma. Para que isso ocorra, é necessário resgatar alguns princípios que definem o SKOS e padrões correlatos, que são as tecnologias semânticas, exequíveis por meio das linguagens de marcação.

Na seção subsequente, são apresentadas as características básicas do XML, como uma linguagem de marcação com princípios descritivos, em contraste com a linguagem HTML, cujo princípio é movido pela apresentação do conteúdo. E sobre o RDF, que surge como um padrão oriundo do XML, mas com propriedades semânticas. A premissa que o define está fundamenta na lógica linguística, por isso, um recurso expresso em RDF pode ser vinculado de tal forma, que seus elementos descritivos obedecem uma ordem semântica prescritiva.

2.3 Tecnologias Semânticas

As novas tecnologias digitais possibilitaram um avanço considerável no campo da Ciência da Informação ao longo das últimas décadas, o que proporcionou o desenvolvimento de instrumentos de representação e organização, provocando convergências entre aportes teóricos e metodológicos de diversas campos científicos, no intuito de atender as demandas informacionais de uma Sociedade cada vez mais baseada em interações virtuais (RAMALHO, 2010).

O resultado dessa convergência orienta a consolidação da web semântica, à medida que os estudos sobre KOS, na Ciência da Informação, têm contribuído na compreensão das estruturas de representação semântica desenvolvidas para o ambiente digital. A web é composta por vários documentos, que muitas vezes não possuem estruturação lógica para conexão semântica entre si. Por isso, a *World Wide Web Consortium* (W3C) tornou-se o órgão regulamentador de boas práticas para publicação de informações e dados no ambiente digital,

e tem contribuído ativamente para uniformização semântica da web, logo, a adoção de padrões reconhecidos pela W3C implica em acessibilidade e interoperabilidade de diversos formatos na web.

Neste sentido, quando se efetua alguma busca pela internet, é possível se deparar com grandes volumes de informações e dados produzidos por seres humanos e processados por máquinas, tal sobrecarga impede o processamento manual desses arquivos. *Browsers* indexam páginas na web diariamente e, para isso, necessitam identificar descritores, palavras-chave, *tags*, termos correlatos ou sentenças para gerar um resultado de busca. O campo da Biblioteconomia e Ciência da Informação tem oferecido subsídios para lidar com os KOS, padrões de metadados e Sistemas de Recuperação de Informações (SRIs), quer sejam automáticos ou manuais, criando formas de representação de recursos informacionais por meio de referências, descrição bibliográfica, resumos, vocabulários controlados, índices, catálogos, entre outros.

A inserção da semântica no ambiente da web, sugere um olhar mais crítico sobre as formas de se representar o conteúdo produzido e disponibilizado, observando padrões interoperáveis e reconhecidos. Neste sentido, as tecnologias semânticas são importantes ferramentas de operacionalização da representação contextualizada no ambiente da web. Sobre isso, Martins e Ramalho (2017) explicam que, as tecnologias semânticas caracterizam-se como linguagens que possibilitam ir além de representações sintáticas, descrevendo, computacionalmente, aspectos semânticos que vão além da estrutura interna e metadados que descrevem os recursos informacionais.

O RDF está no âmago dessa estrutura, pois fundamenta a contextualização descritiva dos recursos informacionais. Mas para entender essa configuração semântica é necessário recorrer a alguns fundamentos que caracterizam o padrão RDF, que são as linguagens de marcação, que possibilitam, inclusive, a criação de modelos de dados com grande capacidade semântica, que no ambiente Web constitui-se como uma das mais ricas fontes de informação (SOUSA; MARTINS; RAMALHO, 2018). Deste modo, apresentam-se os conceitos elementares que caracterizam as linguagens de marcação, especificamente o HTML e o XML, posteriormente se discute sobre o RDF, tendo em vista que é o cerne das tecnologias semânticas e uma das razões para o surgimento do SKOS.

A web como se conhece atualmente, se deve em grande parte a uma linguagem de marcação muito difundida, conhecida como *HyperText Markup Language* (HTML), no entanto,

a prioridade do HTML não era semântica/contextual ou o conteúdo da página em si, porém a apresentação e disposição dos recursos na tela do computador. Logo, é possível que uma página escrita em HTML possua recursos como imagens, vídeos, documentos e sons sem que, necessariamente, haja uma correta descrição ou representação. Não reconhecendo autor, título nem ano, expressando o que por muito tempo foi visto como fragilidade virtual, a respeito de materiais publicados, que não garantia confiabilidade das informações, problema que ainda assombra alguns usuários da *internet*.

A ideia de linguagem de marcação resgata o hábito de marcar um texto, como uma atividade comum na vida de muitos leitores, pois o uso de marcadores sempre está associado a termos-chave, nos quais é possível estabelecer relações entre eles. As linguagens de marcação assemelham-se, de certo modo, a isso, pois sua característica principal é criar marcas (*tags*) para delimitar um texto (FURGERI, 2006). A maioria das páginas da web disponíveis, foram criadas utilizando uma linguagem de marcação, embora a descrição não tenha sido o foco, o HTML tem fornecido o suporte sobre o qual torna possível a navegação na internet, e, ao longo do tempo, tem sofrido várias modificações que culminaram na versão do HTML5, com várias funcionalidades interoperáveis.

A W3C gerencia pacotes de recomendações que padronizam formatos e práticas na disponibilização de recursos e aplicações da web. Isso facilita, sobretudo, a interoperabilidade entre softwares e aplicações para a produção e consumo de informações neste ambiente. Dentre essas diversas recomendações, o XML é um formato de descrição para qualquer tipo de recurso informacional, como o nome propõe, é uma linguagem de marcação extensível. Segundo Moreno e Brascher (2007), o XML é derivado de uma linguagem desenvolvida pela IBM na década de 1980, a *General Markup Language* (GML), uma linguagem de marcação complexa e de alto nível que não se popularizou devido à dificuldade em manipulá-la.

Comparado ao GML, o XML é sua versão simplificada, menos complexa, no entanto extremamente exigente na formatação e validação dos documentos, uma linguagem que não tolera erros em seu escopo. Por ser um formato simples, mas de alto nível qualitativo é amplamente utilizado em soluções web, banco de dados, software de processamento de informações, bibliotecas virtuais e quaisquer mecanismos que necessitam de troca de informações.

De acordo com Lima e Carvalho (2005) o XML é uma metalinguagem, ou seja, ela oferece recursos para a definição de gramáticas que caracterizam linguagens para classes de

documentos específicos, com conjunto de elementos, atributos e regras de composição bem determinados. Isto possibilita descrever qualquer tipo de artefato material ou virtual. O XML permite adicionar qualquer tipo de marcação, por exemplo, os campos de descrição podem ser expandidos para, ano, editora, local, localização, etc. Na figura 3 são apresentadas marcações como <cientistas>, <obra>, <autor> e <titulo>, estas tags não são pré-definidas especificamente, elas podem ser criadas de acordo com a necessidade de quem as elabora.

Figura 3 – Descrição simples de livro em XML

```
<?xml version="1.0" encoding="UTF-8"?>
- <cientistas xsi:noNamespaceSchemaLocation="teste.xsd"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  - <obra>
    <autor>Charles Darwin</autor>
    <titulo>Do telhado das Américas à teoria da evolução</titulo>
  </obra>
</cientistas>
```

Fonte: elaboração própria.

Outra característica do XML são um conjunto de regras expressas no *Document Type Definition* (DTD) ou no *XML Schema*. Entendidas como parâmetros que validam ou confirmam se a estrutura do arquivo XML está correta. O exemplo mostrado na figura 4 apresenta um *XML Schema*, que valida a estrutura de descrição do arquivo XML da figura 3. Para Siqueira (2003) um arquivo DTD é usado para especificar quais elementos, tags ou atributos são permitidos no arquivo XML. Segundo Lima; Carvalho (2005) um documento criado de acordo com uma sintaxe (DTD ou *XML Schema*) é um documento XML válido. Isto não é uma obrigatoriedade para arquivo.xml, todavia ressalta a qualidade dos dados que o compõem.

Figura 4 – Schema XML para validação

```
<?xml version="1.0" encoding="UTF-8"?>
- <xsd:schema elementFormDefault="qualified" xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  - <xsd:element name="cientistas">
    - <xsd:complexType>
      - <xsd:sequence>
        - <xsd:element name="obra" maxOccurs="unbounded">
          - <xsd:complexType>
            - <xsd:sequence>
              - <xsd:element name="autor" type="xsd:string"/>
              - <xsd:element name="titulo" type="xsd:string"/>
            </xsd:sequence>
          </xsd:complexType>
        </xsd:element>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

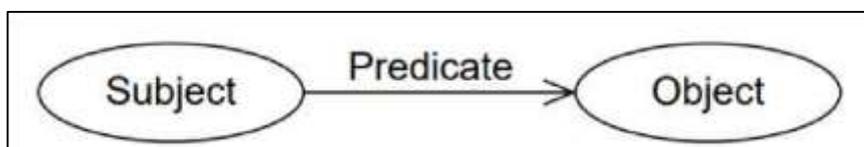
Fonte elaboração própria.

O objetivo de um *XML Schema* é definir uma classe de documentos XML para verificar se um documento XML está em conformidade com um esquema específico

(FALLSIDE, WALMSLEY, 2004). O *XML Schema* apresenta basicamente o que um arquivo XML pode conter, elementos complexos e simples, sua estrutura obedece uma hierarquia. Este é um exemplo de *XML Schema* local, ou seja, está armazenado no mesmo lugar que o arquivo XML, todavia, a maioria dos esquemas estão disponíveis por meio de *Uniform Resource Identifiers* (URIs) públicas para verificar o uso correto dos prefixos utilizados no processo de marcação.

Dentre os diversos padrões e modelos possíveis a partir do XML, o RDF se destaca como forma de descrição e representação de informações no ambiente web. De acordo com Furgeri (2006), o RDF tem como objetivo principal prover intercâmbio de informações entre aplicações sem a perda do significado. Portanto, preocupa-se com a definição de um vocabulário padrão de descrição que possa permitir a interoperabilidade entre aplicações. O RDF funciona por meio de declarativas que expressam a relação tripla entre recurso, conforme a figura 5, que representa o sujeito de uma declaração, propriedade, que é o predicado de uma declaração e valor, o objeto de uma declaração (FURGERI, (2006).

Figura 5 – Tripla RDF



Fonte: Cyganiak; Wood; Lanthaler (2014)

Apesar dos avanços e usos que o envolvem, a perspectiva do RDF na Ciência da Informação é preciso atentar para esta forma de representação no ambiente da web, que atualmente é explorada por diversas instituições, como a *Library of Congress* (LC). De acordo com Ferreira e Santos (2013), o RDF pode utilizar-se do XML como sintaxe comum para o intercâmbio e o processamento de metadados, portanto, a qualidade desse modelo para interoperabilidade de recursos, no ambiente web, está em consonância com a acuidade do *Standard Generalized Markup Language* (SGML). Por ser um modelo reconhecido pela W3C, o RDF estabelece um padrão de metadados para ser embutido na codificação XML (SOUZA; ALVARENGA, 2004).

Ao descrever um arquivo em XML, as *tags* ou marcas, podem ser personalizadas, todavia padrões de metadados como o *Dublin Core* (DC) possui uma série de prefixos na forma de *tags*, que expressam campos de descrição para um recurso informacional. Esses prefixos

padronizados podem ser facilmente importados para um arquivo de descrição cujas características exigem um determinado padrão. Para que isso seja viável, primeiro é necessário coordenar o processo de descrição, isto é possível por meio de esquemas XML que norteiam a prefixação das marcações, estes esquemas são conhecidos por meio de seus *namespaces* que são apontados por uma URI.

Os *namespaces* apontam para esses esquemas XML que permitem associar uma série de padrões, como DC, RDF, SKOS e *Ontology Web Language* (OWL). O uso desses padrões são formas de enriquecer o processo de descrição e vinculação de dados, ou seja, é possível transformar descrições simples em complexas e vice-versa. Isto implica dizer que a possibilidade de desenvolver um modelo semântico personalizado é possível, reunindo características selecionadas de cada padrão. Portanto, é possível estabelecer diversos tipos de relações enquanto um objeto é descrito.

Neste sentido, o RDF oferece subsídios para a construção semântica da web por meio de dados interligados, que podem ser implementados em sistemas de representação e recuperação de informações. De acordo com Martins e Ramalho (2017) atualmente a principal tecnologia que fundamenta a representação semântica na Web é denominada como RDF. E uma das principais características do RDF é a possibilidades de relacionamentos entre recursos. Utilizando não somente a tripla sujeito predicado e objeto, mas o uso expansivo de classes e propriedades, que tornam efetivo esse relacionamento conceitual. A figura 6 apresenta declarações em RDF por meio da referência ao *RDF Schema*, e ao *Dublin Core Schema*, utilizando prefixos reconhecidos ao contexto de uso, elementos estes que torna possível a semântica dos recursos descritos.

Figura 6 – RDF com prefixos DC

```
<?xml version="1.0"?>
<rdf:RDF xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  <rdf:Description rdf:about="http://www.exemplo.org/Darwin">
    <dc:title>Teoria da evolução </dc:title>
    <dc:creator>Charles Darwin </dc:creator>
    <dc:description>Aborda sobre a origem e evolução das espécies no planeta terra</dc:description>
    <dc:language>pt-BR </dc:language>
    <dc:publisher>Editora Brasil </dc:publisher>
    <dc:date>2010 </dc:date>
    <dc:type>Texto </dc:type>
  </rdf:Description>
</rdf:RDF>
```

Fonte: elaboração própria.

As classes e propriedades do RDF podem ser comparadas a conjuntos, no qual é possível relacionar com outros conjuntos, expressando relações hierárquicas ou axiomáticas. Como toda linguagem, entender a sintaxe é saber as regras que condicionam o uso da língua.

No ambiente virtual isso não é diferente, pois trata-se de uma preditiva para leitura dos arquivos em RDF e correta apropriação do assunto.

A relação sujeito, predicado e objeto envolve uma URI, que associada à um termo será como uma espécie de conceito devido ao seu endereçamento único, o predicado refere-se aos prefixos validados nos esquemas indicados. E o objeto apresenta a descrição dos valores associados ao sujeito. Esta tripla do RDF estabelece um vínculo estrito entre o sujeito e o objeto, mediado por predicados que podem ser prefixos de outros *Schemas XML*, indicando qual o tipo de relação estabelecida. Esses atributos em contexto é o que dá semântica ao conteúdo descrito.

Ao usar URIs com nomes para coisas, facilita-se o uso para que as pessoas possam procurar esses nomes e fornecer informações úteis, usando os padrões RDF, ou linguagens de consulta como SPARQL, para buscas por URIs e inclusão de links à outras URIs, para que possam descobrir mais coisas. (BERNERS-LEE, 2006). Isto possibilitará a publicação e recuperação de dados interligados, partindo do pressuposto de que as URIs identificam objetos ou conceitos na web de forma única.

O RDF é uma linguagem de propósito geral para representar informações na web, a *Turtle* é uma sintaxe para RDF, que permite que os gráficos RDF sejam completamente escritos em um formato de texto compacto e natural, com abreviaturas para padrões de uso comuns e tipos de dados. *Turtle* é uma extensão do *N-Triples* compatível com N3. A *Turtle* usa o Identificador de Recursos Internacionalizados (IRIs), bem como URIs, como sinalizar os termos.

Com o RDF, não só é possível descrever, mas apresentar um conteúdo legível por máquinas, pois a partir dos esquemas de referência, os arquivos são validados. O XML, como linguagem padrão, transfere suas propriedades para o RDF e seus conceitos mais elementares precisam ser explorados com propriedade por pesquisadores da Ciência da Informação, pois se trata da lógica de funcionamento dos vocabulários vinculados. O XML e o RDF são utilizados para descrever recursos, e sua arquitetura é o que dá suporte ao SKOS, o uso de regras que classificam e hierarquizam essas informações pertencem intrinsecamente a esse processo. A seguir são apresentadas as principais características dos SKOS, vocabulários que estão codificados nesse padrão e alguns parâmetros de avaliação que podem ser adotados.

3 SIMPLE KNOWLEDGE ORGANIZATION SYSTEM

O SKOS favorece uma aproximação entre instrumentos e métodos de construção de vocabulários controlados utilizados tradicionalmente no campo da Ciência da Informação, com as novas potencialidades oferecidas pelas tecnologias semânticas. Deste modo, destaca-se a necessidade de esforços interdisciplinares, que respeitando as especificidades de cada área do conhecimento, possam contribuir na evolução dos processos de representação e organização do conhecimento de forma integral.

Neste sentido, busca-se identificar, nas teorias e instrumentos da Organização do Conhecimento, a utilização de substratos essenciais que possibilitem uma melhor compreensão, para avaliação e uso do padrão SKOS explorando o seu surgimento, especificações técnicas e formas de avaliação. Esta parte divide-se em duas, na **seção 3.1**, discute-se sobre vocabulários que estão disponíveis no padrão SKOS. Na **seção 3.2**, são apresentados alguns parâmetros que foram identificados na literatura científica para avaliação de vocabulários em SKOS.

Conforme relatam Pastor-Sanchez, Martínez-Mendez, e Rodriguez-Muñoz (2012) o desenvolvimento do SKOS começou em 2002, mas em agosto de 2009 com a publicação do SKOS *Simple Knowledge Organization System Reference* (MILES; BECHHOFFER, 2009a), o SKOS passou a ser considerado como uma recomendação oficial do W3C como modelo de dados para a disponibilização e compartilhamento de KOS na Web. Mais adiante será constatado que a iniciativa para o surgimento do SKOS começou um pouco antes desta data segundo BAKER et al (2013).

As Recomendações SKOS do W3C são projetadas para suportar publicações de vocabulários, como tesouros, na *web*. E na sua essência é um modelo de dados que distingue os conceitos e etiquetas para representá-los. Broughton (2010) ressalta a proximidade entre SKOS e classificações facetadas, ressaltando que algumas aplicações tecnológicas apresentam semelhanças com os tradicionais sistemas de organização facetados, de modo que a teoria da classificação facetada pode fornecer um modelo para os elementos utilizados nos sistemas de representação baseados em tecnologias semânticas, bem como terminologias para habilitá-los.

Mustafa El Hadi (2015) também contextualiza que até o momento as pesquisas relacionadas ao desenvolvimento de novos sistemas de representação têm se focado em questões de interoperabilidade técnica e semântica, porém, existe a falta de um arcabouço teórico geral para considerar estes desenvolvimentos. Neste sentido, Mustafa El Hadi (2015)

afirma que o padrão SKOS inaugura uma nova etapa de sistemas de representação, capazes de proporcionar uma dimensão semântica praticamente inexplorada em ambientes digitais.

O SKOS originou-se do esforço de Phil Cross, Dan Brickley e Traugott Koch para expressar tesouros em RDF, para que isso fosse possível, transformaram os resultados deste esforço em uma proposta, publicada em conjunto pelo *Institute for Learning and Research Technology* (ILRT) no Reino Unido e a *Lund University Library Netlab* na Suécia, para codificar um conjunto de relacionamentos de tesouro usando um esquema RDF (BAKER et al, 2013). Inserido no projeto europeu LIMBER (*Language Independent Metadata Browsing of European Resource*, 1999-2001), foi definido um vocabulário baseado em "conceitos" rotulados por termos em vários idiomas.

BAKER et al (2013), destacam que os resultados do Projeto LIMBER foram incorporados ao projeto SWAD Europa (*Semantic Web Advanced Development*, 2001-2004). Na SWAD Europa, Alistair Miles, do *Rutherford Labs*, solicitou contribuições de especialistas em sinônimos e padrões de classificação. Posteriormente, o vocabulário foi revisado por uma comunidade de usuários interessados e publicado sob o nome de "*Simple Knowledge Organization System*". Em 2004, a *W3C Semantic Web Best Practice and Deployment Working Group* (2004- 2006) cria uma página inicial chamada "*SKOS Core*". Em 2005, o grupo de trabalho publica o "*SKOS Core Vocabulary Specification*" como *W3C Working Draft*. (BAKER et al, 2013).

De acordo com Isaac; Summer (2009), o padrão SKOS possibilita que conceitos possam ser identificados utilizando-se URIs, etiquetados com sequências textuais em uma ou mais línguas naturais, documentados por meio de diferentes tipos de notas, relacionados semanticamente entre si, a partir de hierarquias informais e redes associativas; e, agregados em diferentes esquemas conceituais. Segundo Maculan (2015), o SKOS possui uma extensão XL (*eXtension for Labels*), que possibilita a formalização de declarações em RDF, representando relações que podem ser interpretadas (legíveis) por máquinas, auxiliando na interoperabilidade entre diferentes vocabulários. Pastor-Sánchez, Martínez-Méndez e Rodríguez-Muñoz (2012) complementam também que que SKOS possibilita estabelecer vínculos entre conceitos de diferentes esquemas, favorecendo uma aproximação com as novas tendências de pesquisa computacionais.

Para entender como o padrão SKOS estrutura-se em um vocabulário controlado, é necessário identificar as etiquetas que prefixam as categorias de uso de cada conceito, no quadro

3 são apresentados os prefixos correspondentes a classe e esquemas conceituais, rótulos lexicais, notações, propriedades da documentação, relações semânticas, coleções conceituais e propriedades de mapeamento. Estas categorias de uso estão subentendidas no corpo da explicação sobre cada prefixo com as principais definições encontradas no *SKOS Primer*, *SKOS Reference* e *SKOS-XL Namespace Document - HTML Variant*.

Quadro 03 – Vocabulário SKOS

CATEGORIAS	PREFIXO
Classe Conceitual	<skos:Concept>
Esquemas Conceituais	<skos:ConceptScheme>
	<skos:inScheme>
	<skos:hasTopConcept>
	<skos:topConceptOf>
Rótulos Lexicais	<skos:altLabel>
	<skos:hiddenLabel>
	<skos:prefLabel>
Notações	<skos:notation>
Propriedades de Documentação	<skos:changeNote>
	<skos:definition>
	<skos:editorialNote>
	<skos:example>
	<skos:historyNote>
	<skos:note>
	<skos:scopeNote>
Relações Semânticas	<skos:broader>
	<skos:broaderTransitive>
	<skos:narrower>
	<skos:narrowerTransitive>
	<skos:related>
	<skos:semanticRelation>
Coleções Conceituais	<skos:Collection>
	<skos:OrderedCollection>
	<skos:member>
	<skos:memberList>
Propriedades de Mapeamento	<skos:broadMatch>
	<skos:closeMatch>
	<skos:exactMatch>
	<skos:mappingRelation>
	<skos:narrowMatch>
	<skos:relatedMatch>

Fonte: Adaptado de Isaac; Summers (2009)

O SKOS é uma aplicação do RDF, que permite que conceitos compostos publicados na *web* sejam vinculados e formem esquemas conceituais. Cada conceito é apontado por uma

URI e rotulado por uma *tag* ou *string* em linguagem natural, além disso, fornece uma linguagem de modelagem conceitual leve e intuitiva para desenvolver e compartilhar novos KOS. Além disso, pode ser facilmente estendido para atender a requisitos específicos. Segundo Isaac; Summers (2009), o SKOS introduz a classe *<skos:Concept>*, que permite aos implementadores afirmar que um determinado recurso é um conceito. Isso é feito em duas etapas:

a) Criando (ou reutilizando) um URI para identificar exclusivamente um conceito;

b) Afirmando no RDF, usando a propriedade *<rdf:type>*, que o recurso identificado por esse URI é do tipo *<skos:Concept>*.

O SKOS fornece três propriedades para anexar rótulos de recursos conceituais, são elas: *<skos:prefLabel>*, *<skos:altLabel>* e *<skos:hiddenLabel>*. O *<skos:prefLabel>* possibilita atribuir um rótulo léxico preferido a um recurso. O *<skos:altLabel>* permite atribuir um rótulo léxico alternativo a um conceito. Para Isaac; Summers (2009); o *<skos:hiddenLabel>*, é um rótulo léxico para um recurso, acessível a aplicativos que executam operações de indexação e pesquisa baseadas em texto, mas não gostariam que esse rótulo fosse ser visível de outra forma. Os rótulos ocultos podem, por exemplo, ser usados para incluir variações de palavras com erros ortográficos de outros rótulos lexicais.

As relações semânticas são umas das principais características do SKOS, neste aspecto evoca as categorias fundamentais de relações que são usadas em vocabulários de acordo com a ISO 25964 por meio de três propriedades padrão. Para relações hierárquicas, amplas e estreitas, utiliza-se respectivamente *<skos:broader>* e *<skos:narrower>* para permitir a representação de links hierárquicos, como a relação entre um gênero e suas espécies mais específicas, ou, dependendo de interpretações, a relação entre um todo e suas partes (ISAAC; SUMMERS 2009).

As relações associativas utilizam *<skos:related>*, que permitem a representação de links associativos (não hierárquicos), como a relação entre um tipo de evento e uma categoria de entidades que normalmente participam dele. Outro uso para o *<skos:related>* é entre duas categorias, onde nenhuma é mais geral ou mais específica. (ISAAC; SUMMERS, 2009). O *<skos:related>* permite a representação de links associativos (não hierárquicos), também podem ser usados para representar *links* de parte inteira que não são entendidos como relacionamentos hierárquicos

As notas de escopo são outras propriedades do SKOS que auxiliam o processo de descrição, ela é para fins de documentação, como a propriedade *<skos:note>*, geralmente são inspiradas por parâmetros dos KOS. Esta propriedade pode ser estendida para um nível mais analítico utilizando propriedades *<skos:scopeNote>*, *<skos:definition>*, *<skos:example>*, e *<skos:historyNote>* para conter outros tipos mais específicos de documentação. (ISAAC; SUMMERS, 2009). Além dessas anotações destinadas aos usuários de um esquema conceitual, o SKOS inclui duas especializações *<skos:note>* úteis para gerentes ou editores de KOS: *<skos:editorialNote>* e *<skos:changeNote>*. O quadro 4 apresenta as definições associadas a cada propriedade.

Quadro 4 – SKOS Note e duas definições

Propriedades <skos:note>	Definição
<i><skos:scopeNote></i>	Fornece algumas informações possivelmente parciais, sobre o significado pretendido de um conceito, especialmente como uma indicação de como o uso de um conceito é limitado na prática de indexação.
<i><skos:definition></i>	Fornece uma explicação completa do significado pretendido de um conceito.
<i><skos:example></i>	Fornece um exemplo do uso de um conceito.
<i><skos:historyNote></i>	Descreve mudanças significativas no significado ou na forma de um conceito:
<i><skos:editorialNote></i>	Fornece informações que auxiliam a limpeza administrativa, como lembretes de trabalhos editoriais ainda a serem feitos ou avisos, caso futuras alterações editoriais possam ser feitas:
<i><skos:changeNote></i>	Documenta mudanças refinadas em um conceito, para fins de administração e manutenção:

Fonte: Isaac; Summers (2009)

É importante ressaltar que o vínculo hierárquico entre *<skos:note>* e suas diferentes especializações permite que toda a documentação associada a um conceito seja recuperada de maneira direta. Pois todo *<skos:definition>* é um *<skos:note>*, todo *<skos:scopeNote>* é um *<skos:note>*, logo, todas as propriedades *<skos:note>* compartilham o mesmo efeito de recuperação (ISAAC; SUMMERS, 2009).

Outra propriedade do SKOS é o *<skos:ConceptScheme>*, uma agregação de um ou mais conceitos, uma classe de conceitos que pode ser criada e usada como entidades independentes. Na prática, quando um recurso do esquema de conceito é criado, ele pode ser vinculado aos conceitos que ele contém, usando a propriedade *<skos:inScheme>* (ISAAC;

SUMMERS, 2009). Para fornecer um acesso eficiente aos pontos de entrada de hierarquias conceituais mais amplas e restritas, o SKOS define a propriedade *<skos:hasTopConcept>*, que permite vincular um esquema conceitual aos conceitos mais gerais que ele contém.

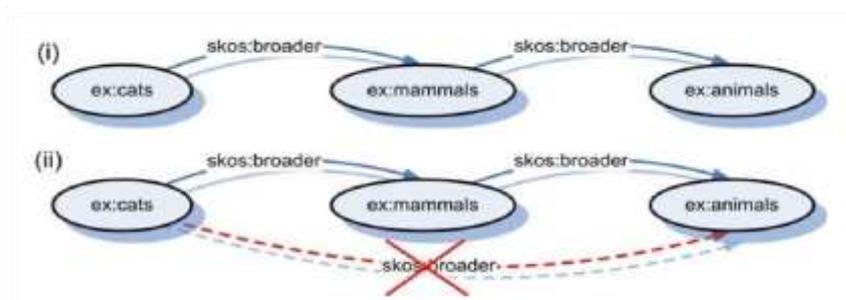
O SKOS fornece propriedades que mapeiam conceitos entre diferentes esquemas de conceito. Isso pode ser feito quando dois conceitos têm um significado semelhante, usando as propriedades *<skos:closeMatch>* e *<skos:exactMatch>*. Dois conceitos de diferentes esquemas de conceito também podem ser mapeados usando propriedades paralelas a estas relações semânticas, que são: *<skos:broadMatch>*, *<skos:narrowMatch>* e *<skos:relatedMatch>*.

Uma propriedade *<skos:closeMatch>* indica que dois conceitos são suficientemente semelhantes para serem usados de forma intercambiável em aplicativos que consideram os dois esquemas de conceito aos quais pertencem. No entanto, o *<skos:closeMatch>* não é definido como transitivo, o que impede que tais avaliações de similaridade se propaguem além desses dois esquemas. O *<skos:exactMatch>* também indica uma semelhança semântica, pois é uma subpropriedade de *<skos:closeMatch>*, no entanto, denota um grau ainda maior de proximidade: os dois conceitos têm significados equivalentes e o link pode ser explorado em uma ampla gama de aplicativos e esquemas. (ISAAC; SUMMERS, 2009).

Uma coleção de conceitos pode ser marcada ou ordenada com SKOS, as coleções marcadas utilizam para modelar corretamente tais estruturas, a classe *<skos:Collection>* e as instâncias de conceitos específicos do grupo de classes, por meio da propriedade *<skos:member>*. Nas coleções ordenadas a classe *<skos:OrderedCollection>* é usada juntamente com a propriedade *<skos:memberList>*. Essa propriedade vincula uma instância *<skos:OrderedCollection>* a um nó do tipo *<rdf:List>*, seguindo o padrão que permite a definição de coleções RDF. (ISAAC; SUMMERS, 2009).

Além de agrupar conceitos, o SKOS permite que inferências sejam aplicadas entre conceitos por meio de hierarquias transitivas. Embora as propriedades usadas para representar hierarquias KOS *<skos:broader>* e *<skos:narrower>* sejam definidas como transitivas (i) e (ii), sua semântica não suporta inferências do tipo: se "animais" é mais amplo que "mamíferos" e "mamíferos" é mais amplo que "gatos", então "animais" é mais amplo que "gatos". A figura 7 ilustra a impossibilidade dessa inferência.

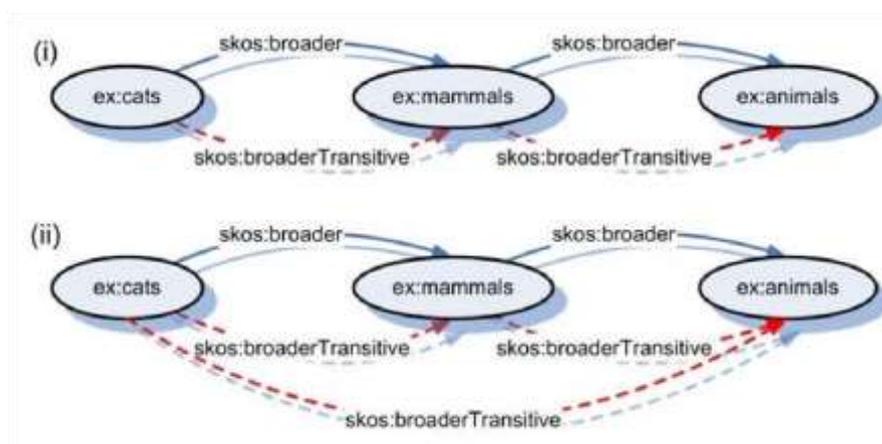
Figura 7 – Propriedades do skos:broader



Fonte: Isaac; Summers (2009)

Por isso o SKOS apresenta duas propriedades específicas: *<skos:broaderTransitive>* e *<skos:narrowerTransitive>*, que são definidas como superpropriedades transitivas de *<skos:broader>* e *<skos:narrower>*. Essa transitividade estabelece relações de dependências entre os conceitos, nos quais é possível inferir que um conceito possui uma lógica relacional com outro. A figura 8 ilustra como essas superpropriedades possibilitam a aplicação de inferências entre os conceitos.

Figura 8 - Propriedades do skos:broaderTransitive



Fonte: Isaac; Summers (2009)

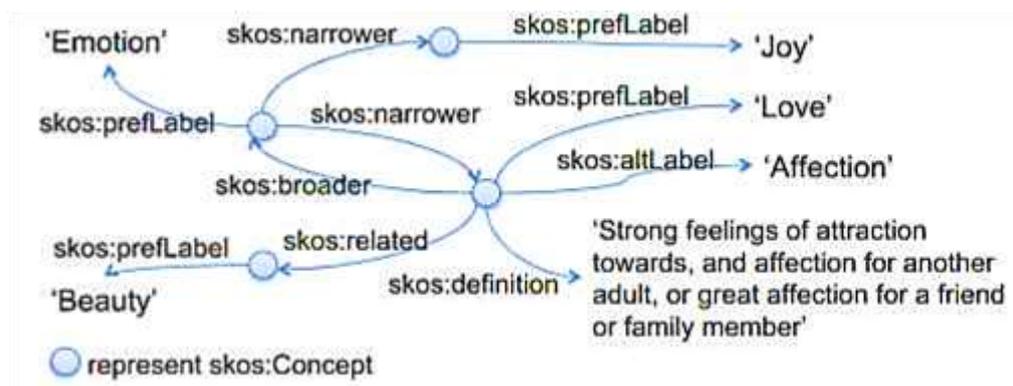
O SKOS é um desses mecanismos que permitem este tipo de manipulação, para contextualizar dados e vincular conceitos, inferindo uma ordem semântica. O que eleva, de forma significativa, o processo de recuperação e compartilhamento da informação na web, o estudo dessas tecnologias implica diretamente na forma como a CI está lidando com estes tipos de tecnologias de representação da informação na web. Diante do contexto emergente, que

envolve toda a sociedade nesse fluxo intermitente de avanços, inovações e terabytes de informações que precisam ser processadas.

Os KOS são agrupamentos conceituais com preditivas lógicas e semânticas aplicadas a recuperação da informação. No aspecto tecnológico, os vocabulários controlados tornaram-se a grande matéria prima para manipulação desses sistemas na web por meios dos dados referentes a objetos. A utilização de vocabulários interligados ajuda na aquisição de conhecimento por controlar rigorosamente e contextualizar os dados, conceitos, objetos, etc. (MÉNDEZ; GREENBERG, 2012).

O SKOS possui uma característica muito comum com os KOS, pois fornece um modelo para expressar a estrutura básica e o conteúdo de esquemas de conceitos. (ISAAC; SUMMERS, 2009). Este modelo de representação de informações na web não tem como objetivo substituir os modelos tradicionais de organização do conhecimento, como vocabulários controlados em seu contexto original de uso, e sim favorecer uma maior reutilização e interoperabilidade entre os vocabulários existentes (RAMALHO, 2015b). Um vocabulário em SKOS contempla alguns aspectos fundamentais, como conceito, termos preferidos e alternativos, termos gerais, específicos e relacionados, a figura 9 apresenta uma visão esquematizada.

Figura 9 – Modelo de vocabulário em SKOS



Fonte: Manaf; Bechhofer; Stevens (2012)

De acordo com Baker et al (2013), a capacidade de rotular um conceito para fins de exibição ou pesquisa é atendida por propriedades para etiquetas preferenciais, alternativas e ocultas (*<skos:prefLabel>*, *<skos:altLabel>* e *<skos:hiddenLabel>*), a adoção dessa nomenclatura equivale aos termos preferidos para uso genérico e específico utilizados em vocabulários controlados, estes termos estão associados à um *<skos:Concept>* apontado por

uma URI, que oferece um caminho único. As etiquetas de *<skos:broader>*, *<skos:narrower>* e *<skos:related>* representam os relacionamentos semânticos entre os termos genéricos, específicos e relacionados.

Portanto, trata-se de um modelo de aplicação de RDF reconhecido como um padrão de representação e compartilhamento de informações na web. O estudo deste padrão possibilita uma reconfiguração das discussões sobre representação da informação na web. Fornecendo insumos para o desenvolvimento e evolução desta temática, no tocante ao desenvolvimento de competências do profissional da informação, diante das novas demandas que surgem devido a influência das tecnologias digitais.

Representar o conteúdo das informações, que são disponibilizadas na web para acesso público é um dos grandes desafios dos modelos semânticos, a busca por padrões que propiciem o processamento semântico por meio de regras lógico-sintática tem oferecido várias alternativas para este ramo. Padrões que atendem esses e outros requisitos, como RDF, SKOS e OWL possibilitam o uso de regras sintáticas e semânticas, envolvendo os recursos informacionais em contextos de descrição.

Estes padrões possibilitam a modelagem conceitual de assuntos por meio de especificações descritivas, para fundamentar este processo, têm se recorrido às técnicas de organização e representação do conhecimento, para convergir fundamentos teóricos e conceituais com modelos lógicos aplicados à tecnologia da informação, principalmente com o uso de tesouros. Gomes e Campos (2004) afirmam, que o tesouro tem sido objeto de interesse da Informática, sendo considerado como um mecanismo para se alcançar a interoperabilidade semântica nos sistemas de representação e recuperação da informação, que favorecem a efetivação de um recurso oferecido.

A conversão de tesouros para vocabulários SKOS, tem se mostrado não somente como algo teórico discutido na literatura científica, mas plenamente exequível, demonstrados, por exemplo, por meio de um método desenvolvido por Assem, Malaisé, Miles e Schreiber, (2006), nos quais apresentam estudos de caso com o *Integrated Public Sector Vocabulary* (IPSV), *GTAA thesaurus* e o *Medical Subject Headings* (MeSH). A conversão desses tesouros operáveis em interoperáveis, por meio de um modelo SKOS, permite a integração de um vocabulário controlado a dados vinculados, pois lhe são atribuídos a tripla semântica que define o padrão RDF.

O SKOS descende da premissa na qual é possível etiquetar termos descritos e, por meio de modelos de relacionamento, atribuir um valor semântico. Segundo MA et al (2011), para promover funções para indexação e navegação de recursos na Web, seria útil codificar os tesouros em formatos compatíveis com a Web. Semelhante ao papel da OWL na edição de ontologias, o SKOS pode ser usado para encadear tesouros, e diversas iniciativas têm comprovado o valor semântico do SKOS, ao ponto de ser considerado como um tipo de ontologia leve, pois sua estruturação se torna menos complexa comparada a uma ontologia declarada em OWL, no entanto, é importante destacar que, nesta pesquisa, uma ontologia é entendida como um vocabulário, assim como um tesouro publicado em SKOS.

Apesar de serem padrões distintos, OWL e SKOS comungam de características provenientes do RDF/XML, o que possibilita tanto a adoção de prefixos do *SKOS Core Vocabulary* em ontologias, como prefixos OWL em tesouros SKOS, ambos para especificar ou estabelecer um vínculo semântico entre os termos descritos. Neste sentido, afirma-se que não é só possível transformar, ou publicar tesouros em SKOS, mas convertê-los em ontologias, migrando de um vocabulário simples para um mais complexo.

Este processo de transição é apresentado por Bechhofer e Miles (2008), que afirmam que, para mover SKOS e OWL em qualquer direção, podemos permitir o padrão de "transformação" sem qualquer impacto em nossa discussão sobre a semântica de *<skos:concept>*. Outra possibilidade é uma hibridização, na qual, o OWL e o SKOS são usados lado a lado para modelar diferentes partes da mesma conceituação (BECHHOFER; MILES, 2008).

Os arranjos terminológicos que estruturam sistemas de recuperação da informação, podem gerar certas limitações, pois dependem da incorporação destes termos e da possibilidade de relacionamento entre eles. As ontologias oferecem suporte a este processo de eficiência, à medida que estruturam logicamente os termos que compõe o universo terminológico de determinada área do conhecimento. O processo de representação consiste na capacidade de projetar um substituto fidedigno do conhecimento, para determinado domínio expresso, por meio da descrição em linguagem específica; em síntese, seria a descrição lógica de tudo o que se sabe sobre determinado objeto e, a partir disso, construir regras de relacionamentos, axiomas e inferências.

A adoção do SKOS na criação de vocabulários é uma iniciativa que oferece alternativas na evolução de um tesouro para uma ontologia, no entanto o simples fato de publicar tesouros em SKOS já possibilita a adoção de *namespaces* que apontem para esquemas

OWL ou qualquer outro tipo. Esse processo de conversão aponta um uso mais amplo para estes modelos semânticos de descrição, pois não é necessariamente uma condição, escolher entre um padrão ou outro, a extensibilidade do XML possibilita entre outros fatores unir propriedades de cada padrão para a criação de novos modelos semânticos que melhor atendam às necessidades de descrição de um recurso informacional.

Apesar destes padrões possuírem características distintas, eles compartilham o mesmo elemento comum, que é o XML, portanto, ainda que seguindo regras hierárquicas para transformar um recurso informacional simples em complexo, como é o caso da conversão de SKOS para OWL, a qualquer momento é possível incluir outros padrões que sigam as mesmas diretrizes do XML como o Dublin Core, configurando-se como um conjunto de padrões interoperáveis.

O SKOS torna possível a tradução da linguagem natural, por meio de elementos descritivos herdados do XML para um formato legível por máquina e simultaneamente característico das linguagens controladas e sistemas de classificação. É importante destacar que, o surgimento de novos padrões, modelos e formatos no ambiente digital, é uma constante realidade, que avança a cada dia, por isso é importante que a Ciência da Informação possa intervir nestas discussões e apresentar suas perspectivas por meio dos estudos de pesquisadores da área. A seguir, são apresentados os principais vocabulários controlados que estão disponíveis em SKOS e foram identificados na literatura.

3.1 Vocabulários controlados em SKOS

O uso de padrões interoperáveis no ambiente da web é uma tendência que tem buscado realizar uma tarefa simples no ambiente tecnológico, trata-se da comunicação eficiente entre máquinas. A Ciência da Informação tem oferecido diversos subsídios teóricos que proporcionam a criação e publicação de vocabulários controlados, e padrões como SKOS surgem como alternativa para alcançar essa interoperabilidade, preservando as características desses insumos informacionais. A identificação de um vocabulário SKOS consiste em verificar a existência de instâncias diretas do tipo *<skos:Concept>*; e instâncias implícitas, devido ao domínio de restrições de alcance nas relações SKOS (MANAF; BECHHOFER; STEVENS, 2012).

Os vocabulários controlados são formas sistêmicas de representar as informações, onde é possível aplicar diversos tipos de relacionamentos lógicos e semânticos. Estruturar a linguagem natural para uma linguagem artificial é um processo que ainda exige muito esforço.

Pois a linguagem continua em transformação e cada língua/idioma sofre mutações que precisam ser revisadas, por isso, vocabulários monolíngues e multilíngues são acompanhados por processos de atualizações.

Não se trata apenas de criar normas, mas entender a variedade conceitual e pragmática dos processos linguísticos e como a tecnologia pode oferecer subsídios para potencializar o desempenho de ferramentas como os vocabulários controlados. A publicação de normas internacionais fornece perenidade no desenvolvimento desta atividade e dá suporte aos processos de inovação que são apontados por tecnologias emergentes.

Com SKOS é possível criar uma série de relações semânticas para estabelecer vínculos entre conceitos diferentes. Ele é composto por termos em um vocabulário denominado *SKOS Core Vocabulary*, que é um conjunto de propriedades e classes utilizadas para expressar o conteúdo e estrutura de um esquema de conceitos em RDF (CATARINO; CERVANTES; ANDRADE, 2015). Os vocabulários controlados estão contidos nos KOS, portanto, como sistema de organização, incorporam a capacidade de representar conceitos por meio de termos significativos que compõem um universo lógico, cujas categorias são classificadas por ordem semântica. Esta qualidade permitiu que a 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).

A norma ISO 25964, que trata sobre tesouros e interoperabilidade com outros vocabulários está dividida em duas partes, a primeira trata sobre tesouros para recuperação de informações, publicada em 2011 abrange os aspectos do desenvolvimento de um vocabulário controlado monolíngue ou multilíngue incluindo um modelo de dados e formatos / protocolos para troca de dados em XML. Esta norma substituiu os padrões anteriores da ISO 2788 sobre Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngues e a ISO 5964 sobre Diretrizes para o estabelecimento e desenvolvimento de tesouros multilíngues. Nesta parte da norma há explicações sobre como configurar mapeamentos entre os conceitos nos vocabulários e outras formas de uso complementar.

A segunda parte trata sobre interoperabilidade com outros vocabulários, publicada em 2013 abrange novas bases não abordadas anteriormente em qualquer padrão. Seu objetivo principal é incentivar a recuperação de informações de alta qualidade em recursos em rede que foram indexados com diferentes vocabulários. Aborda sobre a interoperabilidade entre os tesouros e outros vocabulários, tais como esquemas de classificação, taxonomias e ontologias.

Esta atualização normativa dá ao SKOS uma característica ímpar como modelo padronizado, no qual possui equivalências terminológicas e relacionais com os tesouros. Nesta perspectiva, o SKOS origina-se nos fundamentos dos KOS, concebendo o conceito como unidade basilar do conhecimento (DALBERG, 1978; CATARINO, 2014; LARA, 2015; RAMALHO, 2015b). Portanto, o SKOS projeta uma relação dos KOS no contexto web, observando os padrões escritos em RDF/XML para esquematização conceitual e semântica.

Essas equivalências terminológicas foram convencionadas nos tesouros, ao longo do tempo, e foram aplicadas ao SKOS, por meio dos prefixos <*skos:broader*>, <*skos:narrower*> e <*skos:related*>. Segundo Isaac e Summers (2009), SKOS fornece um modelo para expressar a estrutura básica e o conteúdo de vocabulários controlados, tais como: tesouros, esquemas de classificação, listas de cabeçalho de assunto e taxonomias; de modo que um conceito possa ser formalmente associado a um termo e relacionado com outros conceitos, espelhando as categorias fundamentais de relações usadas em vocabulários. Com intuito de identificar quais vocabulários estão disponíveis na web, no formato SKOS ou XML/RDF, a seguir são apresentados os principais vocabulários citados na literatura e disponíveis atualmente no ambiente da web, no padrão SKOS.

O *Thesaurus multilingual – AGROVOC*, é um vocabulário controlado que abrange todas as áreas de interesse da Organização das Nações Unidas para Agricultura e Alimentação – *Food and Agriculture Organization* (FAO), incluindo alimentação, nutrição, agricultura, silvicultura, pesca, ciência, nomes comuns de animais e plantas, meio ambiente, noções biológicas, técnicas de cultivo de plantas, entre outros. O AGROVOC é um esquema de conceito RDF / SKOS-XL e um conjunto de Dados Abertos Vinculados (LOD) editado pela ferramenta *VocBench*⁴. Os conceitos são agrupados em 25 áreas abrangentes e contém mais de 35.000 conceitos disponíveis em até 29 idiomas. Atualmente, está alinhado com outros 16 vocabulários, o que evidencia sua potencialidade interoperável.

O *General Multilingual Environmental Thesaurus – GEMET*, foi desenvolvido como uma ferramenta de indexação, recuperação e controle para o Centro Europeu de Catálogo de Fontes de Dados (ETC / CDS) e a Agência Europeia do Meio Ambiente (EEA), em Copenhague. O GEMET segue as normas ISO em tesouros monolíngues e multilíngues. Além

⁴ O VocBench é uma plataforma de desenvolvimento colaborativa, multilíngue e baseada na web para o gerenciamento de ontologias OWL, tesouros SKOS (/ XL), léxicos Ontolex-lemon e conjuntos de dados RDF genéricos. Disponível em: <http://vocbench.uniroma2.it/>

disso, possui dois sistemas para organizar os descritores, um esquema de classificação e uma ordem temática, contendo 40 temas.

O *Multilingual Thesaurus of the European Union – EuroVoc* é um tesouro multilíngue e pluridisciplinar sobre as atividades da União Europeia, que contém termos em 23 idiomas. Gerido por um serviço de publicações, avançou para as tecnologias de gestão de sinónimos e web semântica baseadas em ontologias, em conformidade com as recomendações do W3C, assim como as mais recentes tendências nos padrões de tesouros.

O *STW Thesaurus for Economics* é um vocabulário controlado utilizado para representar e pesquisar conteúdo relacionado a economia. Com quase 6.000 cabeçalhos de assuntos em inglês e alemão e mais de 20.000 sinónimos, abrange todas as áreas relacionadas a economia e, em um nível mais amplo, os campos de assuntos relacionados mais importantes. O STW é publicado e continuamente desenvolvido pelo *Leibniz Information Centre for Economics – ZBW*, em acordo com as últimas mudanças na terminologia econômica.

Os *Vocabulários do Getty* são um conjunto composto por 5 vocabulários, o *Art & Architecture Thesaurus (AAT)*, o *Getty Thesaurus of Geographic Names (TGN)*, o *Cultural Objects Name Authority (CONA)*, a *Union List of Artist Names (ULAN)* e o *Getty Iconography Authority (IA)*. Os vocabulários do Getty são construídos para permitir seu uso em dados vinculados e há um projeto para publicá-los na nuvem *Linked Open Data (LOD)*.

O *UNESCO Thesaurus* é uma lista controlada e estruturada de termos utilizados na análise e recuperação de documentos e publicações em diversas áreas. É estruturado em grandes campos de assunto ou domínios, por exemplo, educação, ciência, cultura, ciências sociais e humanas, informação e comunicação, política, direito e economia, países e agrupamentos de países, além disso, divide-se em microtesouros que permitem uma visão geral do assunto.

O *Library of Congress Subject Headings (LCSH)* é mantido desde 1898, para catalogar materiais sob tutela da Biblioteca do Congresso Americano, inclui todos os cabeçalhos de assunto da biblioteca do Congresso, subdivisões flutuantes (tópicas e forma), cabeçalhos de gênero/formulário, cabeçalhos de crianças e cadeias de validação para os registros de autoridade criados. O conteúdo inclui alguns títulos de nomes (pessoais e corporativos) e cabeçalhos geográficos que são adicionados conforme a necessidade de estabelecer subdivisões, ou fornecer um padrão para a prática de subdivisão, ou fornecer estrutura de referência para outros termos.

O **Medical Subject Headings (MeSH)** é um vocabulário controlado produzido pela *National Library of Medicine (NLM)* utilizado na indexação, catalogação e pesquisa de informações de documentos biomédicos e de saúde. Inclui descritores de assunto que aparecem no *MEDLINE/PubMed*, no banco de dados de catálogo do NLM e em outros bancos de dados do NLM, além disso, possui três tipos básicos de Registros: Descritores, Qualificadores e Registros Suplementares de Conceito (SCRs).

O **Agricultural Thesaurus and Glossary (ATG)** foi lançado pela primeira vez pela Biblioteca Nacional Agrícola em 2002. Em 2007, pela *National Agricultural Library, USDA (NAL)* e o Instituto Interamericano de Cooperação para a Agricultura (IICA). Este tesouro é organizado em 17 categorias de assunto, indicadas pela designação "*Subject Category*". As categorias de assunto são usadas para navegar no tesouro em uma disciplina específica ou área de assunto, inclui relações hierárquicas, equivalentes e associativas entre conceitos.

O **Vocabulário Controlado de Governo Eletrônico (VCGE)** é um esquema para ser utilizado no elemento assunto/categoria (*subject.category*) do Padrão de Metadados do Governo Eletrônico brasileiro, está organizado em uma poli-hierarquia. Ou seja, um determinado termo pode ter um ou mais Termos Gerais (TG).

O **Tesouro Brasileiro de Ciência da Informação – (TBCI)** é um vocabulário controlado concluído no final de 2013 e anunciado no Enancib daquele ano, mas somente em 2014 é publicado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Em 2017, deu-se início ao processo de conversão do TBCI do formato PDF para o software gerenciador de tesouros, denominado de TEMATRES⁵, iniciativa desempenhada por José Carlos Francisco dos Santos, Brígida Maria Nogueira Cervantes e Mariângela Spotti Lopes Fujita, do Grupo de Pesquisa “Estudos Terminológicos em Ciência da Informação”, vinculado ao Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina – UEL.

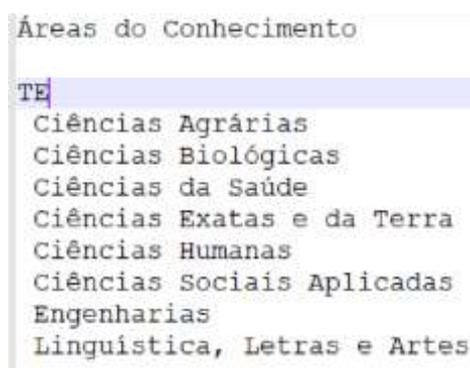
Esses vocabulários são amplamente conhecidos na área de Ciência da Informação e possuem diversos trabalhos e pesquisas relacionadas, no entanto, a avaliação de vocabulários controlados precisa ser constante, pois as publicações no ambiente digital são influenciadas por diversos fatores, entre eles a possibilidade de atualizações periódicas, conversão para novos formatos e formas de processamento. Segundo Mader, Haslhofer e Isaac (2012), a especificação

⁵ TemaTres é um aplicativo web de código aberto para servidor utilizado para gerenciar e explorar vocabulários, tesouros, taxonomias e representações formais de conhecimento.

SKOS define um conjunto de condições de integridade que determinam se dados padrões de dados são consistentes com o modelo SKOS.

Para exemplificar como SKOS se comportam enquanto padrão de vocabulário controlado, converteu-se um trecho do TBCI, especificamente sobre as áreas do conhecimento, que foram classificadas de acordo com CNPq e utilizadas como referência sistemática para o respectivo vocabulário, a figura 10 apresenta essa classificação terminológica.

Figura 10 – Áreas do conhecimento TBCI



Fonte: Adaptado de (IBICT, 2014)

Para realizar esta conversão, utilizou-se o software *TemaTres* para inserção dos termos e indicação dos relacionamentos. A figura 11 apresenta a codificação, em SKOS, de um trecho das áreas do conhecimento apresentado no TBCI, a versão completa pode ser consultada no apêndice. Cada conceito possui uma URI que precisa ser referida de acordo com sua função, quer seja o conceito principal, como é o caso das áreas do conhecimento ou dos conceitos secundários, por exemplo, Ciências Agrárias. Neste caso, as marcações `<skos:hasTopConcept>`, `<skos:ConceptScheme>`, `<skos:Concept>`, `<skos:broader>` e `<skos:narrower>` executam essa função.

Figura 11 – Trecho das áreas do conhecimento do TBCI em SKOS

```

<skos:Concept rdf:about="http://tematestel.000webhostapp.com/vocab/?tema=47">
<skos:prefLabel xml:lang="pt">Áreas do Conhecimento</skos:prefLabel>
<skos:inScheme rdf:resource="http://tematestel.000webhostapp.com/vocab/" />
<skos:narrower rdf:resource="http://tematestel.000webhostapp.com/vocab/?tema=48" />
<skos:narrower rdf:resource="http://tematestel.000webhostapp.com/vocab/?tema=49" />
<skos:narrower rdf:resource="http://tematestel.000webhostapp.com/vocab/?tema=50" />
<skos:narrower rdf:resource="http://tematestel.000webhostapp.com/vocab/?tema=51" />
<skos:narrower rdf:resource="http://tematestel.000webhostapp.com/vocab/?tema=52" />
<skos:narrower rdf:resource="http://tematestel.000webhostapp.com/vocab/?tema=53" />
<skos:narrower rdf:resource="http://tematestel.000webhostapp.com/vocab/?tema=54" />
<skos:narrower rdf:resource="http://tematestel.000webhostapp.com/vocab/?tema=55" />
<dc:created>2018-09-05 00:05:09</dc:created>

</skos:Concept>
<skos:Concept rdf:about="http://tematestel.000webhostapp.com/vocab/?tema=48">
<skos:prefLabel xml:lang="pt">Ciências Agrárias</skos:prefLabel>
<skos:inScheme rdf:resource="http://tematestel.000webhostapp.com/vocab/" />
<skos:broader rdf:resource="http://tematestel.000webhostapp.com/vocab/?tema=47" />
<dc:created>2018-09-05 00:06:34</dc:created>

</skos:Concept>
<skos:Concept rdf:about="http://tematestel.000webhostapp.com/vocab/?tema=49">
<skos:prefLabel xml:lang="pt">Ciências Biológicas</skos:prefLabel>
<skos:inScheme rdf:resource="http://tematestel.000webhostapp.com/vocab/" />
<skos:broader rdf:resource="http://tematestel.000webhostapp.com/vocab/?tema=47" />
<dc:created>2018-09-05 00:06:34</dc:created>

</skos:Concept>
<skos:Concept rdf:about="http://tematestel.000webhostapp.com/vocab/?tema=50">
<skos:prefLabel xml:lang="pt">Ciências da Saúde</skos:prefLabel>
<skos:inScheme rdf:resource="http://tematestel.000webhostapp.com/vocab/" />
<skos:broader rdf:resource="http://tematestel.000webhostapp.com/vocab/?tema=47" />
<dc:created>2018-09-05 00:06:34</dc:created>

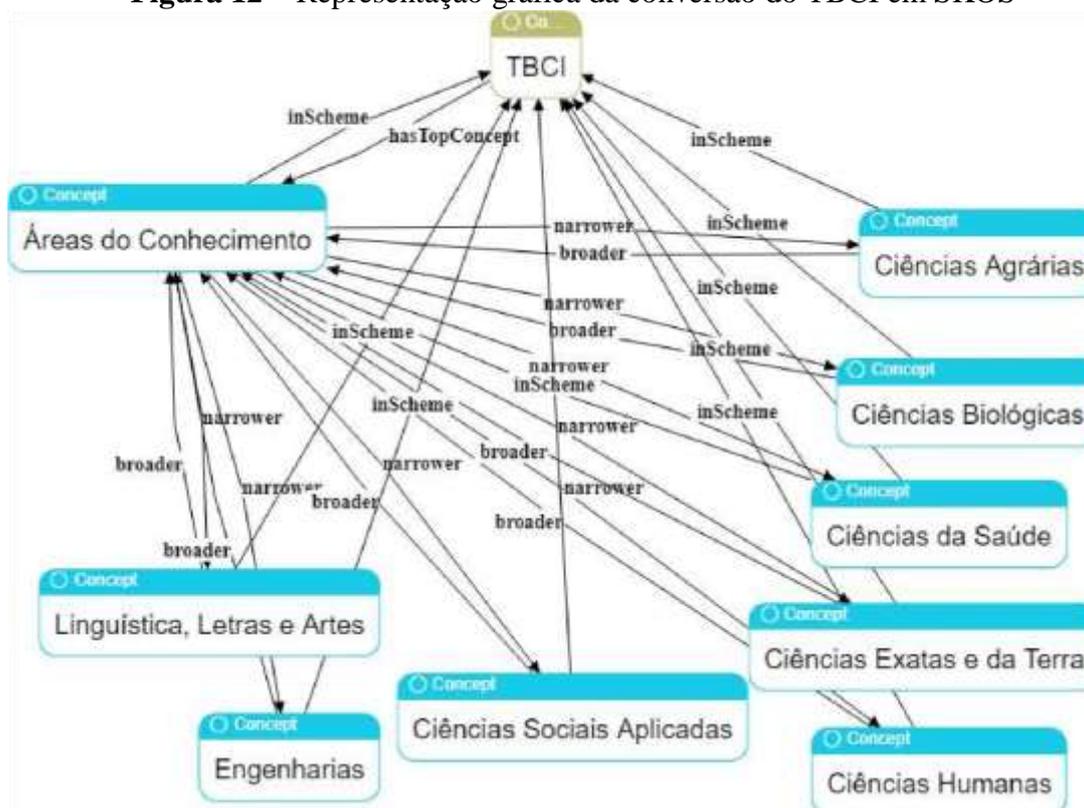
</skos:Concept>

```

Fonte: elaboração própria.

Para representar o conceito principal do esquema de conceito, utiliza-se <skos:hasTopConcept>. O conjunto de termos que se relaciona com o conceito principal é expresso por <skos:ConceptScheme>, que opcionalmente pode incluir declarações sobre relações semânticas entre esses conceitos (MILES; BRICKLEY, 2005). Os conceitos contidos no <skos:hasTopConcept> e <skos:ConceptScheme> são declarados por meio do <skos:Concept> e, para representar em linguagem natural, utiliza-se <skos:prefLabel>. A Figura 12 destaca essa conversão na forma de grafo.

Figura 12 – Representação gráfica da conversão do TBCI em SKOS



Fonte: elaboração própria.

As propriedades semânticas SKOS, diferente de uma ordem hierárquica, tornam os termos relacionais. Ou seja, o mesmo termo pode ser genérico quando inserido em assuntos amplos ou específico se estiver contido em uma categoria mais abrangente. Além da reciprocidade semântica, quando se elege determinado termo que se adequa ao contexto temático, isto é, nesta situação, também é possível estabelecer uma relação mais ampla ou estreita simultaneamente.

SKOS retoma as características dos KOS quando emula os conceitos por meio de URIs, além de associar termos aos tipos de relacionamentos possíveis em vocabulários. Não obstante, a representação pode ser feita com vocabulários simples, como cabeçalhos de assuntos até redes semânticas, que possuem uma estrutura mais complexa. O RDF transfere as propriedades de vínculos entre dados, que é mais conhecido como *Linked Data*, neste sentido, a publicação de vocabulários vinculados é uma característica semântica que acompanha os vocabulários em SKOS.

Mader, Haslhofer e Isaac (2012), ressaltam que a integridade do SKOS não consegue capturar os aspectos de qualidade como quantidade de termos em idiomas diferentes,

conceitos com termos idênticos e sinônimos desconectados. Pois, a principal razão encontra-se na abordagem de “compromisso mínimo” do SKOS. Nesse caso, trata-se de um padrão que visa a interoperabilidade de domínio, e para que isso ocorra deve abster-se de definir restrições que imponha um domínio sobre os requisitos de outro. O SKOS é, portanto, “muito liberal” com a qualidade dos dados (MADER; HASLHOFER; ISAAC, 2012).

Portanto, a avaliação de vocabulários publicados em SKOS é uma necessidade identificada a partir da literatura e da norma ANSI/NISO Z39.19, neste sentido, a identificação de meios para avaliar esses vocabulários é uma tarefa que pode ser contribuída por diversos *cases*, que trazem experiências e resultados de etapas seguidas no processo de avaliação.

Na próxima seção, são apresentados alguns conceitos e tipos de avaliação de vocabulários controlados identificados por meio da National Information Standards Organization (2010), Soergel (2002), Lancaster (2002), Gil-Leiva (2008), Soler-Monreal (2009), Soler-Monreal e Gil-Leiva (2011), Manaf; Bechhofer e Stevens (2012), Mastora, Peponakis e Kapidakis (2017), Mader, Haslhofer e Isaac (2012), Suominen e Mader (2014) e Suominen e Hyvönen (2012). Esses autores, possibilitaram apresentar alguns parâmetros que podem ser utilizados na avaliação do padrão SKOS para a representação de vocabulários controlados.

3.2 Parâmetros para a avaliação do padrão SKOS

A avaliação é uma etapa necessária em todos tipos de sistemas, com o KOS não é diferente. Conforme destaca Mustafa El Hadi (2015), desde a padronização do SKOS, muitos vocabulários passaram a ser protegidos por direitos autorais, para serem publicados como dados ligados na Web, proporcionando uma dimensão semântica praticamente inexplorada anteriormente. Clarke e Zeng (2012) destacam a importância do padrão SKOS como um modelo de dados que distingue expressamente os conceitos e etiquetas representados.

Pastor-Sanchez, Martinez-Mendez e Rodriguez-Muñoz (2009), afirmam que a exploração dos vocabulários na Web é muito limitada, as diferenças na forma de recuperação de informações na Internet exigem vocabulários interoperáveis e multilíngues. Tais autores insistem que o próprio conceito de tesouro tem evoluído para adaptar-se aos modelos de representação de dados na web, abandonando o paradigma lexical em favor de um paradigma conceitual.

Segundo Sanchez-Jiménez e Gil-Urdiciain (2007), essa adaptação permite superar as deficiências das relações fortemente ancoradas ao nível léxico, que caracterizam os tesouros tradicionais. Apesar de existirem diversos manuais para elaboração de vocabulários controlados, os métodos de avaliação dos mesmos têm sido menosprezados, talvez pelo fato de existir um intenso esforço de muitos profissionais e especialistas da área. Desde a proposta até a consolidação do produto final, no entanto, sua avaliação é algo tão necessário quanto sua concepção.

A elaboração de um vocabulário controlado envolve o esforço de muitos profissionais, todavia isto não pode impedir que seja avaliada periodicamente. O aumento na publicação de vocabulários vinculados no ambiente web, revela um grande salto na propagação da interoperabilidade semântica em catálogos online, entretanto, quando não observadas as regras de manutenção e avaliação, um vocabulário pode se tornar obsoleto, por ser um produto que opera com linguagem, portanto, suscetível a variações ao longo do tempo.

A avaliação é uma etapa necessária aos diferentes tipos de KOS, por isso, no âmbito da Ciência da Informação existem distintos tipos de avaliação, que são propostas para atender esta demanda. Incluem-se métodos com abordagens objetivas, subjetivas, automáticas ou manuais. Podendo enfatizar usabilidade, estrutura, relevância e a qualidade geral de um sistema de organização.

Apesar de existirem diversos manuais para elaboração de vocabulários controlados, os métodos de avaliação dos mesmos têm sido menosprezados. Tal afirmação é reforçada, quando percebe os escassos tipos de avaliação para vocabulários controlados. Para embasar esta perspectiva, são apresentados os principais autores e entidade que abordaram sobre avaliação de vocabulários tradicionais no Ciência da Informação (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010; SOERGEL, 2002; LANCASTER, 2002) com indicações mais gerais, (GIL-LEIVA, 2008; SOLER-MONREAL, 2009; SOLER-MONREAL, GIL-LEIVA, 2011), propondo métodos quantitativos e qualitativos.

Quando aplicados aos vocabulários em SKOS, as avaliações recorrem a verificações estruturais que estão especificadas nas recomendações da W3C e as propriedades comuns aos vocabulários vinculados. Neste contexto se destacam Manaf, Bechhofer e Stevens (2012); Mastora, Peponakis e Kapidakis (2017); Mader, Haslhofer e Isaac (2012); Suominen e Mader (2014), e Suominen e Hyvönen (2012), com abordagens de análises automáticas.

Os parâmetros apresentados estão divididos em dois tipos, o primeiro direciona-se para os vocabulários tradicionais e o segundo para vocabulários em SKOS (quadro 5), com os respectivos autores supracitados. Deste modo, explora-se os principais procedimentos adotados para os distintos tipos de vocabulários controlados.

A partir da identificação de algumas referências de avaliação de vocabulários encontrados na literatura científica, foi possível simplificar as propostas desses parâmetros. Quer sejam em infraestruturas tradicionais (impresso ou publicados em padrões incompatíveis com formatos interoperáveis) ou digitalmente interoperável como SKOS.

Quadro 5. Tipos de avaliação de vocabulários

<i>Vocabulário</i>	<i>Autor/Entidade</i>
Tradicional	National Information Standards Organization (2010)
Tradicional	Soergel (2002)
Tradicional	Lancaster (2002)
Tradicional	Gil-Leiva (2008), Soler-Monreal (2009), Soler-Monreal; Gil-Leiva (2011)
SKOS	Manaf; Bechhofer; Stevens (2012)
SKOS	Mastora; Peponakis; Kapidakis (2017)
SKOS	Mader; Haslhofer; Isaac (2012)
SKOS	Suominen; Mader, (2014)
SKOS	Suominen; Hyvönen (2012)

Fonte: elaborado pelo autor

A National Information Standards Organization (2010) destaca que há pouca pesquisa ou literatura que envolva, especificamente, teste e avaliação de vocabulários controlados. O que pode ser devido, em parte, à dificuldade de isolar a qualidade de um vocabulário a partir do desempenho de outros elementos de um sistema de informação, como tecnologia, mecanismo de pesquisa e interface. Esta norma alerta que os seguintes métodos devem ser considerados, quando confrontados com o desafio de medir a qualidade ou eficácia de um vocabulário controlado, referindo-se, portanto, a três tipos de avaliação

- a) A avaliação heurística, realizada por um perito ou um quórum de especialistas, tal avaliação pode ser informal e qualitativa, ou formal e quantitativa.

- b) A modelagem de afinidade, que reúne uma amostra representativa de usuários para avaliar um vocabulário, solicita uma classificação destes termos. Os resultados são analisados contra a hierarquia de termos existentes, e os usuários também podem ser convidados a pontuar termos equivalentes e relacionados de acordo com o nível de similaridade.
- c) E o teste de usabilidade, uma avaliação holística do sistema de informação que pode fornecer informações sobre a eficácia do vocabulário controlado.

Esses tipos de avaliações transmitem aspectos gerais sobre os tipos de avaliação para vocabulário, por isso, a própria ANSI/NISO Z39.19 cita o trabalho de Soergel (2002), que apresenta uma série de critérios que podem ser adotados no processo de avaliação de um vocabulário controlado.

Para Soergel (2002), perguntas podem ser feitas para englobar aspectos sintáticos, semânticos e conceituais dos termos, assim como os seus relacionamentos, podendo conter critérios adicionais aplicáveis aos vocabulários acessíveis eletronicamente, se são hiperlinks ativos de referências cruzadas, se existe um índice de termos online ou pesquisa por palavra-chave, etc. (SOERGEL, 2002). Enquanto isso, Lancaster (2002) defende que algumas medidas podem ser utilizadas no processo de avaliação de um vocabulário, como razão de equivalência, razão de reciprocidade, definição, flexibilidade, nível de pré-coordenação e tamanho do grupo de termos.

Gil-Leiva (2008) e Soler-Monreal (2009), estudando métodos de avaliação de vocabulários do tipo tesouro, apresentam uma divisão em grupos principais, que denominam de avaliação qualitativa intrínseca e avaliação intrínseca quantitativa, ou estatística extrínseca. Para Soler-Monreal e Gil-Leiva (2011), a avaliação intrínseca pode ser realizada com o objetivo de analisar os próprios vocabulários controlados, de modo a estudar sua estrutura, os campos temáticos ou facetas, notas de escopo, relações semânticas, grau de especificidade, etc.

Por outro lado, avaliação extrínseca estuda o impacto nos sistemas de informação, que os utilizam tanto na indexação quanto na recuperação (SOLER-MONREAL; GIL-LEIVA, 2011). Apesar de Soler-Monreal (2009) apresentar esses critérios de avaliação e apresentar o SKOS como linguagem ontológica, não há um direcionamento exato para avaliar esse tipo de vocabulário. Estes são alguns parâmetros que podem ser utilizados para avaliar vocabulários tradicionais.

No que diz respeito a avaliação do padrão SKOS, Manaf, Bechhofer e Stevens (2012), são específicos em sua pesquisa, pois avaliam diretamente vocabulários codificados em SKOS, para compreender os estilos de modelagem usados nesses vocabulários que foram publicados na web, verificando números de conceitos, relacionamentos e suas URIs. O método de avaliação de Manaf; Bechhofer e Stevens (2012) demonstra que, apesar de existir um grande número de vocabulários publicados, muitos não conseguem se adequar totalmente às recomendações da W3C, para estruturação de vocabulários em SKOS.

O trabalho de Manaf, Bechhofer e Stevens (2012) foi realizado em várias etapas, que consistiram em preparar um corpus de vocabulário do candidato SKOS, identificar os vocabulários do SKOS, coletar dados de pesquisa, filtrar várias cópias dos mesmos vocabulários do SKOS e analisar o corpus dos vocabulários.

A preparação do *corpus* incluiu uma série de pesquisas sobre arquivos publicados em skos", ". owl", ".rdf", ".ttl", ".nt", ".n3" e ".xml". Na identificação dos vocabulários, foi adotado um critério que estabelecia que um vocabulário SKOS é um vocabulário que pelo menos contém o (s) conceito (s) SKOS usado (s) diretamente, ou as construções do SKOS que, indiretamente, inferem o uso de um conceito de SKOS, como o uso de relações (MANAF; BECHHOFER; STEVENS, 2012).

A identificação do vocabulário candidato do SKOS consistia em verificar a existência de instâncias diretas do tipo `<skos:Concept>`; e instâncias implícitas de `<skos:Concept>`, devido ao domínio restrições de alcance nas relações SKOS (por exemplo, o assunto de um `<skos:broader>`, `<skos:narrower>` ou `<skos:related>` relacionado necessariamente a um `<skos:Concept>`) (MANAF; BECHHOFER; STEVENS, 2012). Caso não atendesse a esses critérios, o vocabulário era descartado.

Na etapa de análise dos dados coletados, foi levado em consideração o número de conceitos, a profundidade de cada conceito de SKOS e a profundidade máxima da hierarquia de conceito, o número total de links para `<skos:broader>`, `<skos:narrower>` e `<skos:related>`, número total de conceitos *singleton* soltos (conceitos que não estão conectados por quaisquer outros conceitos), o número total de conceitos raiz (conceitos com apenas `<skos:narrower>`, mas não `<skos:broader>`) e o número máximo da propriedade `<skos:broader>` (MANAF; BECHHOFER; STEVENS, 2012).

Na etapa de filtragem de várias cópias dos mesmos vocabulários em SKOS, foi utilizado o registro de informações da etapa anterior, para filtrar vocabulários SKOS

estruturalmente idênticos, comparando o <*ConceptSchemeIRI*> para procurar por vocabulários duplicados. Se os dois vocabulários tivessem registros idênticos, então era verificado o conteúdo, fazendo uma comparação entre os pares das instâncias de <*skos:Concept*> em um vocabulário para o outro. Se os dois vocabulários obtivessem as mesmas instâncias de <*skos:Concept*>, então uma cópia era removida (MANAF; BECHHOFER; STEVENS, 2012).

Quando os dois vocabulários não tinham registros idênticos ou instâncias idênticas de <*skos:Concept*>, era assumido que um vocabulário fosse uma versão mais recente do outro, portanto, era necessário verificar se sua categoria era a mesma como tesouro, taxonomia ou glossário, então, mantinha-se a última versão do vocabulário e removia-se a versão mais antiga. Caso contrário, ambos os vocabulários eram mantidos (MANAF; BECHHOFER; STEVENS, 2002). Na etapa de análise da pesquisa, Manaf, Bechhofer e Stevens (2012) introduziram uma métrica SKOS, por meio de uma fórmula para criar parâmetros de identificação automática que pudessem indicar quando um vocabulário fosse tesouro, taxonomia ou glossário.

De acordo com Manaf, Bechhofer e Stevens (2012), foi coletado um total de 6.819 vocabulários candidatos do SKOS, após o estágio de identificação de vocabulário do SKOS, 1.068 vocabulários foram identificados como vocabulários do SKOS, de acordo com a definição estabelecida, a filtragem de vocabulários SKOS, estruturalmente idênticos, resultou na exclusão de 603 vocabulários idênticos do SKOS e 11 versões mais antigas dos vocabulários do SKOS, que resultou em 454 vocabulários SKOS para uma análise mais aprofundada e, com uma correção dos deslizamentos de modelagem, resultou em 24 vocabulários adicionais do SKOS no *corpus*, que ofereceu aos pesquisadores o número final de 478 vocabulário em SKOS, dos quais, 61% eram taxonomias; 27% glossários, 11% thesaurus e 1% classificado como outros.

Este método de avaliação de Manaf, Bechhofer e Stevens (2012) demonstrou que, apesar de existir um grande número de vocabulários publicados, muitos não conseguem se adequar ao mínimo requerido em sua estruturação, chamando a atenção de vocabulários vinculados, publicados na web, que podem não seguir as recomendações da W3C no seu processo de elaboração.

No estudo realizado por Mastora, Peponakis e Kapidakis (2017), no âmbito de avaliação de vocabulários para identificar até que ponto cada KOS é apropriado para migração para o SKOS, foram selecionados três tipos de KOS, nomeadamente o thesaurus multilingue *Eurovoc*, o *Library of Congress Subject Headings* (LCSH) e o *Dewey Decimal Classification*

(DDC); justificados pelo fato de tesouros, cabeçalhos de assunto e esquemas de classificação, serem os três tipos mais frequentes de KOS usados no contexto das bibliotecas.

A análise desses vocabulários foi realizada por um *parser*, desenvolvido pelo *Greek Institute for Language* para lidar somente com textos gregos. Os vocabulários selecionados possuíam traduções na língua grega, considerada, pelos autores, como a língua natural com menor índice de dupla interpretação. Mastora, Peponakis e Kapidakis (2017) concluem, ao final da análise, que a conversão de dados tradicionais para RDF não é uma condição suficiente para adquirir todo o potencial da Web Semântica. Ainda mais, em casos como os KOS tradicionais, onde há uma certa falta de semântica bem definida, portanto, torna-se necessário realizar algum tipo de reengenharia na construção ou conversão de vocabulários em SKOS.

Dentre os métodos de avaliação apresentados, (SUOMINEN, HYVÖNEN, 2012; MADER, HASLHOFER, ISAAC, 2012; SUOMINEN, MADER, 2014) são os que mais se aproximam de uma análise precisa, porque conseguem aliar fundamentos normativos e métricas da infraestrutura conceitual do SKOS, logo, é o principal parâmetro destacado na avaliação do padrão SKOS, para a representação de vocabulários controlados.

Por meio destes respectivos trabalhos é possível apresentar uma compilação de parâmetros aplicáveis aos vocabulários SKOS. Pois conseguem englobar grande parte dos requisitos essenciais, dispostos nas recomendações W3C e da ISO 25964. Como avaliar um vocabulário não é um procedimento orientado por uma ação comum, Suominen e Hyvönen (2012), Mader, Haslhofer e Isaac (2012) e Suominen e Mader, (2014) identificam erros recorrentes em vocabulários SKOS, que podem ser analisados por meio de verificação automática.

A incidência de tais erros pode comprometer a qualidade de um vocabulário, portanto, indica o que não deve conter em um vocabulário codificado em SKOS. Tais parâmetros endossam o processo de avaliação, porque indicam os passos para adoção de critérios que garantam a qualidade de um vocabulário em SKOS.

Suominen e Hyvönen (2012) avaliaram 14 vocabulários em SKOS e identificaram cerca de 14 erros comuns, categorizados como: URIs válidos, *tags* de idioma ausentes, etiquetas ausentes, conceitos soltos, classes OWL disjuntas, uso consistente de etiquetas, uso consistente de propriedades de mapeamento, uso consistente das relações semânticas, valores ambíguos de etiquetas preferidas, sobreposição disjuntiva, propriedades de etiqueta, relações semânticas

disjuntivas, hierarquia mais ampla em ciclos e espaço extra em branco. (SUOMINEN; HYVÖNEN, 2012).

Mader, Haslhofer e Isaac (2012) avaliaram 15 vocabulários e classificam os problemas recorrentes em vocabulário SKOS em três categorias mais amplas, especificando os problemas contidos em cada uma delas. A primeira categoria é classificada como Rotulagem e Documentação, que inclui: *tags* de idioma omitido ou inválido, cobertura incompleta de idiomas, conceitos não documentados, conflitos de etiqueta.

A segunda categoria refere-se a questões estruturais, que incluem: conceitos órfãos, componentes fracamente conectados, relações hierárquicas cíclicas, relações associativas sem valor, conceitos únicos transitivamente relacionados, conceitos superiores omitidos e conceito superior com conceitos mais amplos (MADER; HASLHOFER; ISAAC, 2012).

A terceira aplica-se aos problemas específicos de dados vinculados, que incluem: links ausentes, links externos ausentes, links quebrados e recursos SKOS indefinidos. Essa categoria refere-se notadamente a problemas de recursos informacionais, que podem estar desvinculados dos seus endereços de localização (MADER; HASLHOFER; ISAAC, 2012).

Suominen e Mader (2014) avaliaram 24 vocabulários e mantiveram as 3 categorias principais de problemas, mas expande sua cobertura, resultando na seguinte classificação:

Etiquetagem e problemas de documentação, abrangendo *tags* de idiomas omitidos ou inválidos, cobertura incompleta de idiomas, conceitos não documentados, etiquetas sobrepostas, etiquetas ausentes, rótulos preferidos inconsistentes, violação de rótulos não contíguos e espaços extras vazios em etiquetas (SUOMINEN; MADER, 2014).

Questões estruturais, incluindo conceitos órfãos, clusters de conceito desconectados, relações hierárquicas cíclicas, relações associativas sem valor, conceitos únicos transitivamente relacionados, conceitos superiores omitidos, conceitos principais não marcados, conceitos principais como conceitos mais amplos, conceitos relacionados unidirecionalmente, confrontos relacionais, mapeando confrontos e violação de classes disjuntas (SUOMINEN; MADER, 2014).

E os problemas específicos de dados vinculados, como links ausentes, links externos ausentes, links quebrados, recursos SKOS indefinidos, violação do esquema HTTP/URI e URIs inválidos (SUOMINEN; MADER, 2014).

Outros problemas foram identificados e incluídos no rol de categorias de análise dos problemas de etiquetagem e documentação, nesta instância foram adicionados: caracteres não imprimíveis em etiquetas, etiquetas vazias, referências de notação ambígua. Nas Questões Estruturais, foram incluídos: redundância hierárquica, relações reflexivas.

As questões estruturais envolvem, diretamente, problemas de ordem lógica, que comprometem o entendimento contextual dos conceitos em um vocabulário, quando se trata de análises automáticas. É uma das principais preocupações, porque tal falha implica diretamente em erro sintático, semântico e conceitual.

Os problemas relativos aos dados vinculados retomam a discussão da manutenção semântica entre os recursos representados, a violação ou quebra de relacionamentos prejudica a contextualização do objeto representado. Para Sousa, Martins e Ramalho (2018), os modelos de dados com capacidade semântica constituem-se como ricas fontes de informação, além de proporcionar a interoperabilidade na troca dessas informações, portanto, sua manutenção é de suma importância.

Deste modo, evidencia-se a necessidade de avaliação de vocabulários controlados, que assumem características de ordem semântica, que, sobretudo, precisam aderir às boas práticas de publicação de vocabulários interligados, assim como os fundamentos que subsidiam essa tecnologia, conforme destaca Martins e Ramalho (2017). A partir dos fundamentos que estruturam os vocabulários controlados, quer sejam normas, parâmetros, padrões ou modelos de representação, é possível identificar pontos de referência que sustentarão análises futuras. O resultado dessa pesquisa trata-se de uma fonte referencial de métodos que pode ser utilizada para analisar diferentes tipos de vocabulários SKOS. Levando em consideração características fundamentais inerentes aos vocabulários controlados.

Diante dos problemas recorrentes na estruturação de um vocabulário que adota SKOS como linguagem codificada de representação, Suominen e Hyvönen (2012), Mader, Haslhofer e Isaac (2012) e Suominen e Mader, (2014) fornecem subsídios para identificação dos problemas mais comuns na avaliação de um vocabulário SKOS, que contém as categorias principais dos erros encontrados em vocabulário SKOS. O quadro 6 apresenta uma adaptação das categorias citadas pelos respectivos autores.

Quadro 6 – Categorização dos erros identificados

CATEGORIA	ERRO
Rotulagem e Problemas de Documentação	Rótulos vazios
	Tags de idioma omitido ou inválido
	Cobertura incompleta de idioma
	Conceitos não documentados
	Não há idiomas comuns
	Rótulos ausentes
	Etiquetas sobrepostas
	Referências de notação ambígua
	Caracteres não imprimíveis em etiquetas
	Conceitos órfãos
	Clusters de conceito desconectados
Questões Estruturais	Relações hierárquicas cíclicas
	Relações associativas sem valor
	Conceitos únicos transitivamente relacionados
	Conceitos superiores omitidos
	Conceitos principais com conceitos mais amplos
	Redundância hierárquica
	Conceitos relacionados unidirecionalmente
	Conceitos relacionados reflexivamente
Problemas específicos de dados vinculados	Falta de links externos
	Recursos SKOS não definidos
	Violação do esquema de URI HTTP

Fonte: elaborado a partir de Mader; Haslhofer; Isaac, 2012, Suominen; Mader (2014)

Uma avaliação que consiga detectar esses tipos de erros, evidencia sua importância, principalmente quando aplicada a vocabulários muito extensos e a validação de conversões para o padrão SKOS. A utilização de métricas automáticas auxilia na objetivação deste processo, pois simplifica os critérios de avaliação orientados pela própria recomendação que especifica o uso do SKOS. Para melhor explorar esses tipos de erros, o quadro 7 apresenta uma definição para cada tipo, como é apresentado a seguir.

Quadro 7 – Definição dos erros

ERRO	DEFINIÇÃO
Rótulos vazios	Rótulos que não contém informações textuais.
Tags de idioma omitido ou inválido	Tags sem idioma especificado
Cobertura incompleta de idioma	Tags parcialmente declaradas em outros idiomas
Conceitos não documentados	Conceitos que não possuem notas de documentação
Não há idiomas comuns	Ausência de idiomas comuns entre os conceitos
Rótulos ausentes	Ausência de rótulos prefLabel, altLabel, rdfs:label
Etiquetas sobrepostas	Dois ou mais conceitos com a mesma etiqueta.
Referências de Notação Ambígua	Conceitos que contém notações skos dentro do mesmo esquema conceitual
Caracteres não imprimíveis em etiquetas	Caracteres que não são alfanuméricos ou espaços em branco nas etiquetas pref / alt / hiddenlabels.
Conceitos órfãos	Conceitos sem relações associativas ou hierárquicas.
Clusters de conceito desconectados	Conceitos aglomerados fracamente conectados ou isolados.
Relações hierárquicas cíclicas	Problema lógico nos relacionamentos mais amplos.
Relações associativas sem valor	Conceitos análogos conectados por uma relação associativa
Conceitos únicos transitivamente relacionados	Uso exclusivo das propriedades skos: broaderTransitive e skos: narrowerTransitive para relacionar hierarquicamente dois conceitos.
Conceitos superiores omitidos	Ausência de pontos de acesso para ConceptScheme.
Conceitos principais como conceitos mais amplos	Subordinação de conceitos principais.
Redundância hierárquica	Redundância nos relacionamentos genéricos e específicos.
Conceitos relacionados unidirecionalmente	Ausência de relações recíprocas usando termos genéricos, específicos, relacionados e conceitos principais.
Conceitos relacionados reflexivamente	Conceitos relacionados aos temas.
Falta de links externos	Links que conectam o vocabulário a outros recursos da Web.
Recursos SKOS não definidos	Uso de termos não autorizados no campo do prefixo SKOS ou elementos SKOS obsoletos.
Violação do esquema de URI HTTP	Uso de outros esquemas de URIs, como URNs e DOIs.

Fonte: elaborado a partir de Mader; Haslhofer; Isaac, 2012, Suominen; Mader (2014)

Os principais problemas de etiquetagem para vocabulários SKOS, referem-se a inconsistências que podem decorrer do processo de inserção ou conversão dos termos, além de possíveis erros de tradução e informações incompletas inseridas como notas de documentação.

Os problemas estruturais podem comprometer a semântica de um vocabulário, pois prejudicará a ordem conceitual de um vocabulário com redundâncias lógicas sem valor. Portanto, a identificação desse tipo de erro reduz a probabilidade de inconsistências de organização conceitual.

Os problemas de dados vinculados aplicados aos vocabulários SKOS, estão diretamente relacionados com a conectividade entre os recursos informacionais representados e o endereço que indica a sua localização. Logo, trata-se de problemas de links quebrados, ausência de direcionamento e URIs inválidas. A identificação desse tipo de erro, aumenta as chances de recuperação de um recurso informacional e da qualidade dos dados vinculados, que são publicados como vocabulários.

Os vocabulários controlados têm demonstrado grande importância enquanto tipo de KOS, pois incorporam fundamentos que facilmente se adaptam em novos suportes tecnológicos. O aumento na publicação de vocabulários vinculados no ambiente da web, revela um grande salto na propagação da interoperabilidade semântica entre KOS. Entretanto, quando não se observa as regras de construção, manutenção e avaliação, um vocabulário pode se tornar problemático, do ponto de vista técnico e conceitual.

A qualidade de um vocabulário controlado exige a análise de requisitos específicos dos recursos oferecidos, como infraestrutura, disponibilidade, usabilidade, entre outros fatores, todavia percebe-se uma dificuldade de avaliação de vocabulários de forma holística. Portanto, a apresentação dos parâmetros de avaliação do padrão SKOS para vocabulários, destacados neste trabalho, oferecem um passo que indica um possível avanço dessa temática no contexto da Ciência da Informação.

4 CONSIDERAÇÕES FINAIS

Desde a criação da ISKO, a Ciência da Informação tem demonstrado grande interesse nos estudos da Organização e Representação do Conhecimento (ORC), as reflexões filosóficas, e até epistêmicas, têm direcionado as discussões sobre a essência da representação e sua percepção no âmbito dos KOS, portanto é possível afirmar, de acordo com o progresso desta pesquisa, alguns pontos que merecem destaque.

O primeiro ponto a ser destacado refere-se ao surgimento da própria Ciência da Informação, que possui uma relação intrínseca com a tecnologia. Essa relação é uma das formas de instrumentalizar suas próprias teorias, como foi a princípio a ideia do MEMEX, desde então, os caminhos para imersão teórica têm direcionado os esforços de diversos pesquisadores da área, com o intuito de fortalecer suas teorias, tendo em vista a carência identificada por (HJØRLAND, 1998).

Esta carência teórica não se configura como uma deficiência simplória, mas uma oportunidade de revisar a *epistême* da área e descobrir novos campos de estudos, como foi o caso da ORC, que evolui a partir da Teoria do Conceito. Kuhn (1998) explica, que as anomalias, ou violações de expectativa atraem a crescente atenção de uma comunidade científica, esta pode ser uma das explicações para ascendência de novos campos de estudos.

As crises na ciência possuem um poder tão transformador, que não se limitam a consternação dos problemas estabelecidos, mas à possibilidade de mudar de um estado para outro a ponto de surgir um novo campo, uma disciplina ou uma ciência completamente distinta das atuais. É esta possibilidade de mudança que tem permitido que áreas como a Organização do Conhecimento reflitam sobre suas teorias e adote uma postura filosófica.

Neste sentido, Deleuze e Guattari (1996, p.13) afirmam que os filósofos, “não devem mais contentar-se em aceitar os conceitos que lhes são dados, para somente limpá-los e fazê-los reluzir, mas é necessário que eles comecem por fabricá-los, criá-los, afirmá-los, persuadindo os homens a utilizá-los. [...]”. Muito mais que adornar uma teoria, é necessário torná-la digna de crítica e reflexão. Além de contestar seu uso, apesar de Deleuze e Guattari (1996) oferecerem essa reflexão crítica que vai de encontro ao pensamento cartesiano, isso não implica dizer que se deve utilizá-lo apenas para contrapor um pensamento cartesiano ou kantiano.

Para Deleuze e Guattari (1996), criar conceitos novos é o objeto da Filosofia, o que não quer dizer que devemos anular os conceitos criados até então, pois considerando o devir equações universais, chegaria um momento em que alguém poderia questionar todos os novos conceitos criados e sugerir uma outra mudança radical e assim sucessivamente. O que Deleuze e Guattari (1996) tentam explicar é que “é inútil perguntar se Descartes tinha ou não razão”, porque os conceitos apresentados por ele responderam as questões a que foram submetidos.

Há que se refletir sobre estes conceitos e a forma como eles foram concebidos, apesar de na Ciência da Informação, logo, a dedicação de Deleuze e Guattari (1996) em explicar o que é um conceito, o plano de imanência, os personagens conceituais, functivos e conceitos, prospectos e conceitos, percepto, afecto e conceito; não pode ser suplantada apenas por uma palavra, mas explorada de tal forma, que possa se repensar a epistemologia da ciência que se propõe a estudar seus fundamentos basilares. Neste sentido, os pesquisadores da Organização do Conhecimento, precisam se questionar sobre o que estão fazendo, por que estão fazendo desta forma e como isso afetará o futuro da Ciência da Informação.

Deleuze e Guattari (1996) oferecem suas contribuições que podem finalizar esta pesquisa a respeito do pensamento crítico sobre os conceitos, enquanto Schopenhauer (2005) apresenta uma profícua discussão sobre a origem dos conceitos, que muitas vezes passa despercebida nos estudos da Ciência da Informação. Apesar de Dahlberg (1978) ratificar um pensamento aristotélico, tendo em vista o uso de benefícios da Lógica para explicar os relacionamentos entre conceitos, constata-se que essa disciplina é o elemento comum entre Schopenhauer (2005) e Dahlberg (1978), para subsidiar as discussões sobre a própria origem e propriedade dos conceitos.

Em Wittgenstein (1968) também é possível encontrar a Lógica como elemento intrínseco aos problemas de linguagem, tal característica evidencia a propriedade lógica dos conceitos e os processos que os envolvem. Neste sentido, os KOS não só representam essa característica, como a expandem para o uso semântico, quando empregam seu uso em vocabulários controlados. No ambiente digital essa propriedade é preservada, pois a codificação computacional também obedece uma ordem lógica, assim como todos os elementos processados por uma máquina.

Como a necessidade de revisar os fundamentos dos KOS é constante, o estudo das linguagens que codificam os conceitos é outro fator que ganha evidência. Principalmente no

que diz respeito aos impactos e sua adaptação às tecnologias emergentes. O vocabulário controlado é um tipo de KOS que consegue integrar essa dualidade entre teoria e aplicação, à medida que proporciona a concepção de instrumentos de representação, que podem representar conceitos, estruturas lógicas e semânticas.

O surgimento de tecnologias que incorporam padrões e modelos de representação de dados na web, possui um background amparado por estes fundamentos, como é o caso do SKOS, e que muitas vezes passam despercebidos por quem usufrui de determinada tecnologia. A publicação de um vocabulário controlado no ambiente da web, indica um modo de organização sistêmica e representativa de recursos informacionais, o SKOS é o principal padrão que tem proporcionado isso, porque prioriza a simplificação. Todavia, seu uso não deve ser irrefletido, pois existem condições prescritas nos documentos oficiais deste padrão, que precisam ser seguidas.

A avaliação de vocabulários é um procedimento que pode auxiliar esse processo, porque estabelece parâmetros que validam ou identificam problemas na estruturação do vocabulário, para que isto ocorra é necessário recorrer aos aspectos teóricos e aplicados. Os parâmetros apresentados nesta pesquisa oferecem um ponto de reflexão para considerar os vocabulários controlados atualmente disponíveis, levando a questionar os processos teóricos e práticos que envolvem os constructos de Organização e Representação do Conhecimento.

O SKOS é um padrão que incorpora as premissas básicas de um KOS e permite ser avaliado automaticamente, por isso o conhecimento de sua estrutura merece destaque, porque implica diretamente na forma como um vocabulário controlado será representado. Por isso, os fundamentos utilizados para entender um pouco mais sobre a linguagem XML e o padrão RDF ganham notoriedade, principalmente na perspectiva *Linked Data*. Apesar de ser muito conhecido no ambiente tecnológico, por programadores e desenvolvedores no caso do XML, esse conhecimento muitas vezes passa a ser técnico e aplicado.

Discutir metalinguagens e modelos de descrição no ambiente digital tem sido uma forma de aproximar esses aspectos técnicos às discussões conceituais do fluxo da informação na web. As implicações que envolvem o RDF estão justamente na compreensão técnica de alguns termos, que tradicionalmente advém da Ciência da Computação, todavia isso não restringe seu uso, pelo contrário, projetos como o Dublin Core e MARC/XML, recuperam termos usuais do processo de catalogação.

O SKOS simplifica a publicação de vocabulários controlados, por isso é importante prezar pela qualidade dos vocabulários que adotam esta codificação. Percebe-se que a dificuldade de avaliar vocabulários tradicionais ainda é presente no âmbito da Organização do Conhecimento, todavia, parâmetros de verificação automática propostos em trabalhos como de Suominen e Hyvönen (2012), Mader, Haslhofer e Isaac, 2012 e Suominen e Mader (2014), auxiliam na sistematização de critérios que possam conduzir este processo.

Trata-se, portanto, da tentativa de objetivar e tornar imparcial grande parte da avaliação, para que não atenda a determinados intensões subjetivas, mas que busque a qualidade e a padronização básica de sua estrutura. Neste sentido, os vocabulários controlados em SKOS representam um avanço, porque devido a sua arquitetura é possível criar regras lógicas que identificarão se os requisitos dispostos em sua documentação técnica serão atendidos.

A revisão dos documentos publicados pela W3C e *Dublin Core Initiative*, subsidiou a construção dos exemplos validados para esta pesquisa. Esta pesquisa resulta na exploração do padrão SKOS por meio de vocabulários controlados, nos quais foi possível identificar uma possível relação entre os fundamentos teóricos, que subsidiam os KOS, e os instrumentos tecnológicos utilizados na representação de recursos informacionais. Como pesquisa exploratória, ressalta-se que possui tripla finalidade, que consistem em desenvolver hipóteses, aumentar a familiaridade do pesquisador com um ambiente, fato ou fenômeno, para a realização de uma pesquisa futura mais precisa ou modificar e clarificar conceitos (MARCONI; LAKATOS, 2003).

Portanto, espera-se que esta pesquisa possa contribuir para as discussões sobre representação por meio de vocabulários controlados e tesouros, com uso de padrões interoperáveis para o ambiente da web. Fomentando outras pesquisas que possam expandir essa temática na Ciência da Informação, para análises reflexivas sobre as formas de representação e tecnologias disponíveis que possam realçar a apropriação de insumos informacionais. O SKOS é resultado de um esforço conjunto para tornar viável o uso de tesouros, em contextos como o *Linked Data*, onde é possível obter, de forma eficaz e coesa, a recuperação de informações que se multiplicam no ambiente digital, por isso é importante zelar pela sua qualidade.

Para ressaltar a importância da avaliação de vocabulários publicados no ambiente digital, Nayak et al (2019) explica que os vocabulários modernos compreendem milhões de conceitos e relações, isto quer dizer que houve um aumento exponencial de vocabulários publicados no ambiente da web, principalmente na forma de dados vinculados. Neste sentido,

avaliação de vocabulários que contém milhares de termos e conceitos associados a eles torna-se um desafio que precisa ser enfrentado.

De acordo com Nayak et al (2019), há uma necessidade de técnicas automatizadas, para avaliar a evolução de hierarquias de conhecimento muito grandes que capturam a subsunção semântica lógica. Isto indica que trabalhos futuros podem não apenas abarcar este tema, como expandi-lo, tendo em vista que há pouca exploração da temática que envolve vocabulários controlados, assim como algum tipo de técnica ou metodologia que possa orientar esse processo, quer seja de forma automática ou não.

Portanto, os trabalhos futuros com essa temática serão de amplo interesse para a comunidade científica e se mostrará como uma continuidade desta pesquisa, que pode alcançar outros patamares, dada a relevância e a familiaridade do tema para o campo da Organização da Conhecimento. Por fim, espera-se que este contributo possa integrar o alicerce de novas pesquisas e orientar trabalhos relacionados.

REFERÊNCIAS

- ALBERTONI, Riccardo. LusTRE: a framework of linked environmental thesauri for metadata management. **Earth Science Informatics**, Springer-Verlag GmbH Germany, part of Springer Nature, 2018. Disponível em: <https://doi.org/10.1007/s12145-018-0344-8>. Acesso em: 09 ago. 2018.
- ALMEIDA, Gladis Maria de Barcellos. O percurso da terminologia: de atividade prática à consolidação de uma disciplina autônoma. **Revista Estudos Linguísticos**, n32, 2003. Disponível em: <http://www.gel.org.br/estudoslinguisticos/volumes/32/htm/mesaredo/mr004.htm>. Acesso em: 09 fev. 2018.
- ALVITE DÍEZ, M. L.; et al. Propuesta de representación del tesouro EuroVoc en SKOS para su integración en sistemas de información jurídica. **Scire: Representación y Organización del Conocimiento**, n.16, v.2, p.47-51, 2010.
- ASSEM, Mark van; MALAISÉ, V´eronique; MILES, Alistair; SCHREIBER, Guus. A Method to Convert Thesauri to SKOS. Y. Sure and J. Domingue (Eds.): ESWC 2006, **Springer-Verlag**, Berlin Heidelberg, LNCS 4011, pp. 95–109, 2006.
- AUSTIN, Derek. **Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngues**. Brasília: IBICT, 1993.
- BACHELARD, Gaston. **A epistemologia**. Lisboa: Edições 70 LDA, 2006.
- BAKER, Thomas et al. Key Choices in the Design of Simple Knowledge Organization System (SKOS). **Journal of Web Semantics**, v.20, 2013. Disponível em: <http://www.cs.vu.nl/~guus/papers/Baker13a.pdf>. Acesso em: 05 fev. 2018
- BARBOSA, Alice Príncipe. **Teoria e prática dos sistemas de classificação bibliográfica**. Rio de Janeiro: Instituto Brasileiro de Bibliografia e Documentação, 1969.
- BECHHOFER, Sean; MILES, Alistair. **Using OWL and SKOS**. W3C, 2008. Disponível em: <https://www.w3.org/2006/07/SWD/SKOS/skos-and-owl/master.html>. Acesso em: 05 fev. 2018.
- BERKELEY, George. **Obras filosóficas**. Tradução, apresentação e notas Jaimir Conte. – São Paulo: Editora UNESP, 2010.
- BERNERS-LEE, T. **Linked Data**. 2006. Disponível em: <http://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 24 out. 2018.
- BRÄSCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento? In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. **Anais [...]**. São Paulo: Ancib, 2008.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. **Guia de Interoperabilidade: Manual do Gestor**. Brasília: MPOG, 2012. Disponível em:

https://www.governodigital.gov.br/documentos-earquivos/Guia_de_Interoperabilidade_Manual_do_Gestor_2012.pdf. Acesso em: 18 ago. 2018.

BROUGHTON, Vanda. Concepts and Terms in the Faceted Classification: the Case of UDC. **Knowledge Organization**, v. 37, n. 4, p.270-279, 2010.

BUSH, V. As we may think. **Atlantic Monthly**, v. 176, n. 1, p. 101-108, 1945. Disponível em: <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>. Acesso em: 10 jun. 2018.

CABRÉ, M. T. **La terminología**: Representación y comunicación. Uma proposta de base comunicativa y otros artículos. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. 1999.

CAFÉ, Lúcia Maria de M. de; BRASCHER, Marisa B. M.; SUJII, Mauro K. Elaboração de tesouros utilizando-se o programa de elaboração de tesouros em microcomputador. **Revista de Biblioteconomia, Brasília**, v. 18, n. 2, p. 185-192, jun./dez. 1990.

CATARINO, Maria Elisabete. Simple Knowledge Organization System: construindo sistemas de organização do conhecimento no contexto da Web Semântica. **Informação & Tecnologia (ITEC)**: Marília/João Pessoa, v.1, n.1, p. 17-28, jan./jun., 2014.

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

CATARINO, Maria Elisabete; SOUZA, Terezinha Batista de. A representação descritiva no contexto da web semântica. **TransInformação**, Campinas, v.24, n.2, p.77-90, maio/ago, 2012.

CAVALCANTI, C. R. **Indexação e tesouro: metodologia e técnica**. Brasília: ABDF, 1978. 89p.

CERVO, A. L.; BERVIAN, P. A. **Metodologia científica**. 5. ed. São Paulo: Prentice Hall, 2003.

CLARKE, S. G. D.; ZENG, M.L. **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.

COSTA, M. A. F.; COSTA, M. de F. B. **Metodologia da pesquisa**: conceitos e técnicas. Rio de Janeiro: Interciência, 2001.

CYGANIAK, Richard; WOOD, David, LANTHALER, Markus. **RDF 1.1 Concepts and Abstract Syntax**. W3C, 2014. Disponível em: <http://www.w3.org/TR/rdf11-concepts/>. Acesso em: 27 out. 2018.

DAHLBERG, Ingetraut. **Definitionen aus dem Begriffsfeld Wissensorganisation**. 2006. Disponível em: https://isko-de.org/beitraege/dahlbergdefinition/?fbclid=IwAR1_spFatgTeAfgUX6pPIpW_xlh20_UI7Met oE7SG_28Evaabc3CnbCn4A. Acesso em: 27 out. 2018.

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

DELEUZE, Gilles & GUATTARI, Félix. **O que é a Filosofia?** Tradução de Bento Prado Jr. e Alberto Alonso Muñoz. Rio de Janeiro: Ed. 34, 1996.

DESCARTES, René. **Discurso do método**. Tradução de Maria Ermantina Galvão. São Paulo: Martins Fontes, 2001.

FALLSIDE, David C.; WALMSLEY, Priscilla. **XML Schema Part 0: Primer Second Edition**. 2004. Disponível em: <http://www.w3.org/TR/2004/REC-xmlschema-0-20041028/>. Acesso em: 12 abr. 2018.

FERREIRA, Ana Carolina; MACULAN, Benildes Coura Moreira dos Santos; NAVES, Madalena Martins Lopes. Ranganathan e a teoria da classificação facetada. **Transinformação**, Campinas, v. 29, n.3, p.279-295, 2017. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010337862017000300279&lng=en&nrm=iso. Acesso em: 31 jun. 2018.

FERREIRA, Jaidir Andrade; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. O modelo de dados Resource Description Framework (RDF) e o seu papel na descrição de recursos. **Inf. & Soc.:Est., João Pessoa**, v.23, n.2, p. 13-23, mai/ago. 2013.

FREY, Katie; ACCOMAZZI, Alberto. **The unified astronomy thesaurus: semantic metadata for astronomy and astrophysics**. arXiv, 2018.

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

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GIL-LEIVA, I. **Manual de indización: teoría y práctica**. Gijón: Trea, 2008.

GOMES, Hagar Espanha; CAMPOS, Maria Luiza de Almeida. Tesouro e normalização terminológica: o termo como base para intercâmbio de informações. **DataGramZero** - Revista de Ciência da Informação, v.5 n.6 dez., 2004.

GONZÁLEZ DE GÓMEZ, Maria Nélide. Metodologia de pesquisa no campo da Ciência da Informação. **DataGramZero**: Revista de Ciência da Informação, v.1 n.6 dez., 2000. Disponível em: http://datagramazero.org.br/dez00/Art_03.htm. Acesso em 25 ago. 2018.

GRAY, Alasdair J.G. et al. Finding the right term: Retrieving and exploring semantic concepts in astronomical vocabularies. **Information Processing & Management**, v. 46, n. 4, p. 470-478, 2010. Disponível em: <https://doi.org/10.1016/j.ipm.2009.09.004>. Acesso em: 25 ago. 2018.

GUARINO, N. **Formal Ontology and Information Systems**. [Vilnius]: Vilnius University Faculty of Mathematics and Informatics, 1998.

GUIMARÃES, José Augusto Chaves. **Organização do conhecimento: passado, presente e futuro em um contexto de diversidade cultural**. In: _____. José Augusto Chaves Guimarães; Vera Dodebei. (Org.). Organização do conhecimento e diversidade cultural. 1ed. Marília: ISKO-Brasil; FUNDEPE, v. 1, p. 13-19, 2015.

HAACK, Susan. **Filosofia das lógicas**. Tradução Cezar Augusto Mortari, Luiz Henrique de Araújo Dutra. São Paulo: Editora UNESP, 2002.

HESSEN, J. **Teoria do Conhecimento**. Tradução – João Vergílio Gallerani Cuter. São Paulo: Martins Fontes. 2003.

HJØRLAND, Birger. **Knowledge organization**. In: Hjørland, Birger, ed. ISKO Encyclopedia of Knowledge Organization, v.43, n. 6: 475-84, 2016. Disponível em: http://www.isko.org/cyclo/knowledge_organization. Acesso em: 18 nov 2018.

HJØRLAND, Birger. Theories are knowledge organizing systems (KOS). **Knowledge Organization**, v.42, n.2, p.113-128, 2015.

HJØRLAND, Birger. Theory and metatheory of information Science a new interpretation. **Journal of Documentation**, v. 54, n 5, 1998. Disponível em: https://www.researchgate.net/publication/228717437_Theory_and_metatheory_of_information_science_A_new_interpretation. Acesso em 18 nov 2018. Acesso em: 18 nov 2018.

HODGE, G. **Systems of knowledge organization for digital libraries: beyond traditional authority files**. Washington, D.C.: The Digital Library Federation Council on Library Information Resources, 2000. Disponível em: <http://old.diglib.org/pubs/dlf090/dlf090.pdf>. Acesso em: dez. 2018.

ISAAC, A.; SUMMERS, Ed. (Ed.). **SKOS Simple Knowledge Organization System Primer**: W3C Working Group Note, 18 August 2009. Disponível em: <https://www.w3.org/TR/2009/>. Acesso em: 19 abr. 2018.

ISO 25964. **Thesauri and interoperability with other vocabularies**. Part 1: Thesauri for information retrieval. Geneve: International Standard Organization, 2011.

ISO 25964. **Thesauri and interoperability with other vocabularies**. Part 2: Thesauri for information retrieval. Geneve: International Standard Organization, 2013.

JONQUET, Clément et al. AgroPortal: A vocabulary and ontology repository for agronomy. **Computers and Electronics in Agriculture**, v. 144, p. 126-143, 2018. Disponível em: <https://doi.org/10.1016/j.compag.2017.10.012>. Acesso em: 12 out. 2018.

KÖCHE, J. C. **Fundamentos de metodologia científica**: teoria da ciência e iniciação à pesquisa. Petrópolis, RJ: Vozes, 2002.

KUHN, Thomas. S. **A estrutura das revoluções científicas**. São Paulo: Perspectiva, 1970.

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

LANCASTER, F.W. **El control del vocabulario en la recuperación de la información**. 2ª ed. Valencia: Universidad de Valencia. 2002.

LARA, Marilda Lopes Ginez de. Propostas de tipologias de KOS: uma análise das referências de formas dominantes de organização do conhecimento. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 20, n. esp. 1, p. 89-107, 2015.

LE COADIC, Yves-François. **A ciência da informação**. 2 ed. Brasília: Briquet de Lemos, 2004.

LIMA Júnio César de; CARVALHO, Cedric Luiz de. **Extensible Markup Language (XML)**. Relatório Técnico, Instituto de Informática da Universidade Federal de Goiás, 2005. Disponível em: www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_00205.pdf. Acesso em 27 abr 2018.

LIMA, G. A. B. O. **Organização do conhecimento: pesquisa e desenvolvimento**. In: José Augusto Chaves Guimarães; Vera Dodebei. (Org.). **Organização do conhecimento e diversidade cultural**. 1ed. Marília: ISKO-Brasil; FUNDEPE, v. 1, p. 670-687, 2015.

LOJKINE, J. **A Revolução Informacional**. São Paulo: Cortez, 1995.

MA, Xiaogang et al. A SKOS-based multilingual thesaurus of geological time scale for interoperability of online geological maps”. **Computers & Geosciences**, v. 37, n. 10, p. 1602-1615, 2011. Disponível em: <https://doi.org/10.1016/j.cageo.2011.02.011>. Acesso em 25 ago 2018.

MACULAN, B. C. M. S. **Estudo e aplicação de metodologia para reengenharia de tesauro**: remodelagem do THESAGRO. 2015. 339f. Tese (Doutorado em Ciência da

Informação) - Escola de Ciência da Informação da Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

MADER, Christian; HASLHOFER, Bernhard; ISAAC, Antoine. **Finding Quality Issues in SKOS Vocabularies**. TPDFL, 2012.

MANAF, N. A. A.; BECHHOFER, Sean; STEVENS, Robert. **The Current State of SKOS Vocabularies on the Web**. In.: E. Simperl et al. (Eds.): ESWC 2012, LNCS 7295, SpringerVerlag Berlin Heidelberg, p. 270–284, 2012.

MANN, Thomas. **O Pensamento vivo de Schopenhauer**. Tradução de Pedro Ferraz do Amaral, São Paulo: Livraria Martins Editora, 1955.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Fundamentos de metodologia científica**. 5. ed. São Paulo: Atlas, 2003

MARTINS, Paulo; RAMALHO, Rogério Aparecido Sá. Tecnologias semânticas: fundamentos para o entendimento do conceito de Linked Data. In: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO, 7., 2017, Londrina. **Anais [...]**. Londrina: UEL, 2017. p.1040- 1051. Disponível em: <http://www.uel.br/eventos/cinf/index.php/secin2017/secin2107/paper/view/483/329>. Acesso em 27 out 2018.

MASTORA, A.; PEONAKIS, M.; KAPIDAKIS, S. SKOS concepts and natural language concepts: An analysis of latent relationships in KOSs. **Journal of Information Science**, v. 43, n. 4, p. 492–508, 2017.

MAZZOCCHI, Fulvio. **Knowledge organization system (KOS)**. Knowledge Organization 45, no.1: 54-78. Also available in ISKO Encyclopedia of Knowledge Organization, eds. Birger Hjørland and Claudio Gnoli, 2018. Disponível em: <http://www.isko.org/cyclo/kos>. Acesso em: 20 jun 2018.

MÉNDEZ, Eva.; GREENBERG, Jane. Linked data for open vocabularies and HIVE's global framework. **El Profesional de la Información**, v.21, n.3, p.236-244, 2012.

MILES, A.; BECHHOFER, S. **SKOS Simple Knowledge Organization System Reference**. W3C Recommendation, 2009a. Disponível em: <http://www.w3.org/TR/skos-reference/>. Acesso em: 20 jun 2018.

MILES, Alistair; BECHHOFER, Sean. **SKOS Simple Knowledge Organization System eXtension for Labels (SKOS-XL) Namespace Document - HTML Variant**. 18th August 2009b. Disponível em: <http://www.w3.org/TR/skos-reference/skos-xl.html>. Acesso em: 20 jun 2018.

MILES, Alistair; BRICKLEY, Dan. **SKOS- Core Guide**. W3C Working Draft 2, 2005. Disponível em: <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102>. Acesso em 08 dez 2018.

MORENO, Fernanda Passini; BRASCHER, Marisa. **MARC, MARCXML E FRBR: relações encontradas na literatura**. *Inf. & Soc.:Est.*, João Pessoa, v.17, n.3, p.13-25, set./dez. 2007.

MOTTA, Dilza Fonseca da. **Método relacional como nova abordagem para construção de tesouros**. Rio de Janeiro: SENAI, 1987. 89 p.

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

NATIONAL INFORMATION STANDARDS ORGANIZATION. **Guidelines for the construction, format, and management of monolingual controlled vocabularies**. Baltimore: NISO, 2010. 172 p. (ANSI/NISO Z39.19-2005 R2010).

NAYAK, G., Dutta et al. Automated assessment of knowledge hierarchy evolution: comparing directed acyclic graphs. *Inf Retrieval J*, v.22, p.256-284, 2019. Disponível em: <https://doi-org.ez31.periodicos.capes.gov.br/10.1007/s10791-018-9345-y>. Acesso em: 6 jul. 2019.

NISO. **Understanding metadata**. Bethesda (EUA): National Information Standards Organization, 2004. Disponível em: https://www.lter.uaf.edu/metadata_files/UnderstandingMetadata.pdf. Acesso em: 18 ago. 2018.

NKOS. **Taxonomy of Knowledge Organization Sources/Systems.**, 2000 (revised July 31, 2000). Disponível em: http://nkos.slis.kent.edu/KOS_taxonomy.htm. Acesso em: 18 ago. 2018.

OHLY, H. Peter. **Ingetraut Dahlberg**. ISKO, Encyclopedia of Knowledge Organization, 2018. Disponível em: <https://www.isko.org/cyclo/dahlberg>. Acesso em 04 jun. 2018.

OLIVEIRA, Marlene de. **Origens e Evolução da Ciência da Informação**. In: _____. OLIVEIRA, Marlene de (Coord.). *Ciência da Informação e Biblioteconomia: novos conteúdos e espaços de atuação*. Belo Horizonte: Editora UFMG, 2005.

PASTOR-SANCHEZ, J. A.; MARTINEZ-MENDEZ, F.J.; RODRIGUEZ-MUNOZ, J.V. 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, 2012.

PASTOR-SANCHEZ, Juan-Antonio; MARTÍNEZ-MENDEZ, Francisco Javier; RODRÍGUEZ-MUÑOZ, José Vicente. Advantages of thesauri representation with the Simple Knowledge Organization System (SKOS) compared with other proposed alternatives. **Information Research**, v. 14, 2009. Disponível em: <http://www.informationr.net/ir/14-4/paper422.html>. Acesso em 04 jun. 2018.

PINHEIRO, Lena Vania Ribeiro; LOUREIRO, José Mauro Matheus. Traçados e limites da ciência da informação. **Ciência da Informação**, v. 24, n. 1, 1995.

RAMALHO, R. A. S. **Desenvolvimento e utilização de ontologias em bibliotecas digitais: uma proposta de aplicação**. 2010. 145f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

RAMALHO, R.A.S. Análise do Modelo de Dados SKOS: Sistema de Organização do Conhecimento Simples para a Web. **Informação & Tecnologia (Itec)**, v. 2, p. 66-79, 2015a.

RAMALHO, R. A. S. **Ontologias e Simple Knowledge Organization System (SKOS): aproximações e diferenças**. In: José Augusto Chaves Guimarães; Vera Dodebei. (Org.). **Organização do conhecimento e diversidade cultural**. 1ed. Marília: ISKO-Brasil; FUNDEPE, v. 1, p. 100-107, 2015b.

RAMALHO, R. A. S.; VIDOTTI, S.A.B.G; FUJITA, M.S.L. Web semântica: uma investigação sob o olhar da Ciência da Informação. **DataGramZero - Revista de Ciência da Informação**, v.8, n.6, dez. 2007.

RANGANATHAN, S. R. **Colon Classification**. Second edition. Madras: Madras Library Association, London: E. Goldston, Ltd, 1939.

SANCHEZ-JIMÉNEZ, R; GIL-URDICIÁIN, B. Lenguajes documentales y ontologias. In: **El profesional de la información**. v. 16, n. 6, p. 551-560, 2007.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n.1, p. 41-62, jan./jun. 1996.

SCHANDL, Thomas; BLUMAUER, Andreas. **PoolParty: Using SKOS Thesaurus Management Linked Data**. L. In: _____.Aroyo et al. (Eds.): **ESWC 2010, Parte II, LNCS 6089**, Berlin: Springer-Verlag Heidelberg, p. 421-425, 2010.

SCHOPENHAUER, Arthur. **O mundo como vontade e como representação, I tomo**. Tradução, apresentação, notas e índices de Jair Barboza, São Paulo: Editora UNESP, 2005.

SELL, Carlos **Eduardo**. **Sociologia clássica: Durkheim, Weber e Marx**. Itajaí, 2001.

SHANNON, C.E.; WEAVER, W. **The Mathematical theory of communication**. Urbana: University of Illinois Press, 1949.

SILVA, Jonathas Luiz Carvalho; FREIRE, Gustavo Henrique de Araújo. Um olhar sobre a origem da Ciência da Informação: indícios embrionários para sua caracterização identitária **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, v. 17, n. 33, p. 1-29, jan./abr, 2012. Disponível em: 10.5007/1518-2924.2012v17n33p1. Acesso em 18 nov 2018.

SIQUEIRA, Marcos Antonio. **XML na Ciência da Informação**: uma análise do MARC 21. Dissertação (mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências da Universidade Estadual Paulista – UNESP, Marília, 2003.

SOERGEL, D. **Thesauri and Ontologies in Digital Libraries**: Tutorial. In: Evaluation of thesauri. Joint Conference on Digital Libraries, Portland, Oregon, July 14, 2002. Disponível em: <http://www.dsoergel.com/cv/B63.pdf>. Acesso em 08 set 2018.

SOLER-MONREAL, C; GIL-LEIVA, I. Evaluation of controlled vocabularies by inter-indexer consistency. **Information Research**: An International Electronic Journal, v. 16, n. 4, 2011. Disponível em: <http://www.informationr.net/ir/16-4/paper502.html>. Acesso em 10 jan 2019.

SOLER-MONREAL, M. C. **Evaluación de vocabularios controlados en la indización de documentos mediante índices de consistencia entre indizadores**. Dpto. De Comunicación Audiovisual, Documentación e Historia del Arte Tesis (Doctorado). Universitat Politècnica de València. Valencia, 2009.

SOUSA, Janailton Lopes; MARTINS, Paulo George Miranda; RAMALHO, Rogerio Aparecido Sá. Modelos de representação semântica na era do Big Data. **Brazilian Journal of Information Studies: Research Trends**. v.12, n.3, p.34-p.40, 2018.

SOUTO, L. F. Recuperação de informação em bases de dados: uso de tesouros. **Transinformação**, Campinas, v. 15, n. 1, p. 73-81, jan./abr, 2003.

SOUZA, Renato Rocha; ALVARENGA, Lídia A Web Semântica e suas contribuições para a ciência da informação. **Ci. Inf., Brasília**, v. 33, n. 1, p. 132-141, jan./abr, 2004.

SUOMINEN, O. et al. **Publishing SKOS Vocabularies with Skosmos**. Manuscript submitted for review, 2015.

SUOMINEN, O.; HYVÖNEN, E. Improving the quality of skos vocabularies with skosify. In: International Conference on Knowledge Engineering and Knowledge Management, 18th, 2012, Ireland. **Proceedings** [...]. Ireland, Springer-Verlag, 2012.

SUOMINEN, Osmo; MADER, Christian. Assessing and Improving the Quality of SKOS Vocabularies. **Journal on Data Semantics**, v. 3 n. 1, p. 47-73, 2014.

TÁLAMO, M. F. G. M.; SMIT, J. W. Ciência da informação: pensamento informacional e integração disciplinar. **Brazilian Journal of Information Science**, v. 1, n. 1, 2007. Disponível em: 10.5016/brajis.v1i1.30. Acesso em 20 jun 2018.

TRIVIÑOS, Augusto Nivaldo Silva. **Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação**. São Paulo: Atlas, 1987.

WITTGENSTEIN, Ludwig. **Tractatus Logico-Philosophicus**. Tradução e apresentação de José Arthur Giannotti. São Paulo: Companhia Editora Nacional, 1968.

WÜSTER, E. **Introducción ala teoria general de la terminologia y a la lexicografia terminológica**. Trad. Anne-Cécile Nokerman, Barcelona, IULA, 1998.

ZAFALON, Z. R. Scan for Marc: sintaxe e semântica de registros bibliográficos na conversão De dados analógicos para o formato MAR21 Bibliográfico. **CRB-8 Digital**, São Paulo, v. 6, n. 1, p. 20-32, 2013. Disponível em: <http://revista.crb8.org.br>. Acesso em 20 jun 2018.

ZAPILKO, Benjamin et al. **TheSoz: A SKOS Representation of the Thesaurus for the Social Sciences**. arXiv, 2012.

ZENG, M. L. Knowledge organization systems (KOS). **Knowledge Organization**. International journal devoted to concept theory, classification, indexing, and knowledge representation, Frankfurt, v. 35, n. 2-3, p. 160-182, 2008.

APÊNDICE

ÁREAS DO CONHECIMENTO DO TBCI CONVERTIDO EM SKOS

```

<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:skos="http://www.w3.org/2004/02/skos/core#"
    xmlns:map="http://www.w3c.rl.ac.uk/2003/11/21-skos-mapping#"
    xmlns:dct="http://purl.org/dc/terms/"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
  <skos:ConceptScheme
    rdf:about="http://temateste1.000webhostapp.com/vocab/">
<dc:title>TBCI</dc:title>
<dc:creator>Lena Vania Ribeiro Pinheiro e Helena Dodd Ferrez</dc:creator>
<dc:contributor>Janailton Lopes Sousa, Rogério Aparecido Sá Ramalho</dc:contributor>
<dc:publisher>IBCT</dc:publisher>
<dc:subject></dc:subject>
<dc:description>O Tesouro Brasileiro de Ciência da Informação (TBCI) é um tipo de
vocabulário controlado publicado em 2014 pelo IBICT</dc:description>
<dc:date>2018-09-04</dc:date>
<dct:modified>2018-11-28 16:41:46</dct:modified>
<dc:language>pt</dc:language>

  <skos:hasTopConcept
rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=47"/>
  <skos:hasTopConcept
rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=59"/>
  <skos:hasTopConcept
rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=39"/>
  <skos:hasTopConcept rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=1"/>
  <skos:hasTopConcept
rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=22"/>
  <skos:hasTopConcept
rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=109"/>
  <skos:hasTopConcept
rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=29"/>
  <skos:hasTopConcept
rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=16"/>
  <skos:hasTopConcept
rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=33"/>
</skos:ConceptScheme>

  <skos:Concept rdf:about="http://temateste1.000webhostapp.com/vocab/?tema=47">
<skos:prefLabel xml:lang="pt">Áreas do Conhecimento</skos:prefLabel>
<skos:inScheme rdf:resource="http://temateste1.000webhostapp.com/vocab/">
<skos:narrower rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=48"/>
<skos:narrower rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=49"/>
<skos:narrower rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=50"/>
<skos:narrower rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=51"/>
<skos:narrower rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=52"/>
<skos:narrower rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=53"/>
<skos:narrower rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=54"/>
<skos:narrower
    rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=55"/>
<dct:created>2018-09-05 00:05:09</dct:created>

```

```

</skos:Concept>
<skos:Concept rdf:about="http://temateste1.000webhostapp.com/vocab/?tema=48">
<skos:prefLabel xml:lang="pt">Ciências Agrárias</skos:prefLabel>
<skos:inScheme rdf:resource="http://temateste1.000webhostapp.com/vocab/"/>
<skos:broader rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=47"/>
< dct:created>2018-09-05 00:06:34</dct:created>
</skos:Concept>

<skos:Concept rdf:about="http://temateste1.000webhostapp.com/vocab/?tema=49">
<skos:prefLabel xml:lang="pt">Ciências Biológicas</skos:prefLabel>
<skos:inScheme rdf:resource="http://temateste1.000webhostapp.com/vocab/"/>
<skos:broader rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=47"/>
< dct:created>2018-09-05 00:06:34</dct:created>
</skos:Concept>

<skos:Concept rdf:about="http://temateste1.000webhostapp.com/vocab/?tema=50">
<skos:prefLabel xml:lang="pt">Ciências da Saúde</skos:prefLabel>
<skos:inScheme rdf:resource="http://temateste1.000webhostapp.com/vocab/"/>
<skos:broader rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=47"/>
< dct:created>2018-09-05 00:06:34</dct:created>
</skos:Concept>

<skos:Concept rdf:about="http://temateste1.000webhostapp.com/vocab/?tema=51">
<skos:prefLabel xml:lang="pt">Ciências Exatas e da Terra</skos:prefLabel>
<skos:inScheme rdf:resource="http://temateste1.000webhostapp.com/vocab/"/>
<skos:broader rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=47"/>
< dct:created>2018-09-05 00:06:35</dct:created>
</skos:Concept>

<skos:Concept rdf:about="http://temateste1.000webhostapp.com/vocab/?tema=52">
<skos:prefLabel xml:lang="pt">Ciências Humanas</skos:prefLabel>
<skos:inScheme rdf:resource="http://temateste1.000webhostapp.com/vocab/"/>
<skos:broader rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=47"/>
< dct:created>2018-09-05 00:06:35</dct:created>
</skos:Concept>

<skos:Concept rdf:about="http://temateste1.000webhostapp.com/vocab/?tema=53">
<skos:prefLabel xml:lang="pt">Ciências Sociais Aplicadas</skos:prefLabel>
<skos:inScheme rdf:resource="http://temateste1.000webhostapp.com/vocab/"/>
<skos:broader rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=47"/>
< dct:created>2018-09-05 00:06:35</dct:created>
</skos:Concept>

<skos:Concept rdf:about="http://temateste1.000webhostapp.com/vocab/?tema=54">
<skos:prefLabel xml:lang="pt">Engenharias</skos:prefLabel>
<skos:inScheme rdf:resource="http://temateste1.000webhostapp.com/vocab/"/>
<skos:broader rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=47"/>
< dct:created>2018-09-05 00:06:35</dct:created>
</skos:Concept>

<skos:Concept rdf:about="http://temateste1.000webhostapp.com/vocab/?tema=55">
<skos:prefLabel xml:lang="pt">Linguística, Letras e Artes</skos:prefLabel>
<skos:inScheme rdf:resource="http://temateste1.000webhostapp.com/vocab/"/>

```

```
<skos:broader      rdf:resource="http://temateste1.000webhostapp.com/vocab/?tema=47"/>  
< dct:created>2018-09-05 00:06:36</dct:created>  
</skos:Concept>  
</rdf:RDF>
```