

Linked data inter-vocabulários: uso do modelo SKOS para a ligação de dados conceituais

Raidan Cruz Silveira

Universidade Estadual Paulista, Departamento de Ciência da Informação, Marília, SP, Brasil
ORCID: <https://orcid.org/0000-0002-8976-3391>
raidan.silveira@unesp.br

Daniel Martínez-Ávila

Universidad de León, Área de Biblioteconomía y Documentación, León, España
ORCID: <https://orcid.org/0000-0003-2236-553X>
dmarta@unileon.es

DOI: <https://doi.org/10.26512/rici.v18.n1.2025.56616>

Recebido/Recibido/Received: 2024-11-10

Aceito/Aceptado/Accepted: 2024-12-20

Publicado/Publicado/Published: 2025-03-21

Resumo

O trabalho discute o modelo Simple Knowledge Organization System (SKOS) como tecnologia que integra iniciativas de Linked Data (LD) a partir da representação de três conceitos e suas relações. É uma pesquisa de natureza aplicada, teórica, exploratória e descritiva e abordagem qualitativa. Como método foram utilizados: o programa XML Copy Editor para criar a base sintática em linguagem computacional junto aos elementos do modelo SKOS e sua combinação com outros recursos tecnológicos, o W3C RDF Validator para validar operações descritas em SKOS e o programa Diagrams.net para criar as representações gráficas. Foram representados em SKOS os conceitos: “tecnologias da informação e comunicação” do *Thesaurus Brasileiro de Ciências da Informação* (TBCI) baseado em Ramalho (2017), “Tecnologia da informação” do Catálogo de Autoridades de Assuntos da Biblioteca Nacional (Brasil) e “Tecnologia da Informação” da Lista de Títulos de Assuntos da Biblioteca do Congresso. A conexão entre os conceitos demonstra o potencial do modelo SKOS como integrante do LD, pois estabelece relações entre conceitos exatos e semelhantes de vocabulários distintos em sistemas de informação. Os conceitos representados em SKOS sugerem características para LD pois dispõem de: linguagens que operam com outros sistemas, uso de URIs para localização de conceitos e formato de descrições em outros padrões tecnológicos que sugerem interoperabilidade. Pesquisas futuras podem explorar melhor o SKOS em ambientes de LD, orientados principalmente ao Library Linked Data (LLD) no âmbito brasileiro e levando em consideração iniciativas internacionais mais avançadas.

Palavras-Chave: Dados conectados. Dados conectados a biblioteca. Sistemas de Organização do Conhecimento. Simple Knowledge Organization System.

Linked Data Inter-Vocabularies: Using the SKOS Model for Linking Conceptual Data

Abstract

This paper discusses the Simple Knowledge Organization System (SKOS) model as a technology that integrates Linked Data (LD) initiatives through the representation of three concepts and their relationships. It is applied, theoretical, exploratory, and descriptive research with a qualitative approach. The methods used include: the XML Copy Editor program to create the syntactic base in computational language along with the SKOS model elements and its combination with other technological resources; the W3C RDF Validator to validate operations described in SKOS; and the Diagrams.net program to create graphical representations. The following concepts were represented in SKOS: "information and communication technologies" from the Brazilian Thesaurus of Information Science (TBCI) based on Ramalho (2017), "Information Technology" from the National Library's Authority Catalog (Brazil), and "Information Technology" from the Library of Congress Subject Headings. The connection between the

concepts demonstrates the potential of the SKOS model as part of LD, as it establishes relationships between exact and similar concepts from different vocabularies in information systems. The concepts represented in SKOS suggest features for LD as they provide: languages that interact with other systems, use of URIs for concept location, and description formats in other technological standards that promote interoperability. Future research may further explore SKOS in LD environments, particularly in the context of Library Linked Data (LLD) within Brazil, while considering more advanced international initiatives.

Keywords: Linked data. Library linked data. Knowledge organization systems. Simple Knowledge Organization System.

Linked Data Inter-Vocabulários: Uso del modelo SKOS para la conexión de datos conceptuales

Resumen

Este trabajo discute el modelo Simple Knowledge Organization System (SKOS) como una tecnología que integra iniciativas de Linked Data (LD) a partir de la representación de tres conceptos y sus relaciones. Es una investigación de naturaleza aplicada, teórica, exploratoria y descriptiva con un enfoque cualitativo. Como métodos se utilizaron: el programa XML Copy Editor para crear la base sintáctica en lenguaje computacional junto con los elementos del modelo SKOS y su combinación con otros recursos tecnológicos; el W3C RDF Validator para validar las operaciones descritas en SKOS; y el programa Diagrams.net para crear las representaciones gráficas. Se representaron en SKOS los conceptos: “tecnologías de la información y comunicación” del Tesouro Brasileiro de Ciências de la Información (TBCI) basado en Ramalho (2017), “Tecnología de la información” del Catálogo de Autoridades de Asuntos de la Biblioteca Nacional (Brasil) y “Tecnología de la Información” de la Lista de Títulos de Asuntos de la Biblioteca del Congreso. La conexión entre los conceptos demuestra el potencial del modelo SKOS como parte de LD, ya que establece relaciones entre conceptos exactos y similares de vocabularios distintos en sistemas de información. Los conceptos representados en SKOS sugieren características para LD, ya que disponen de: lenguajes que operan con otros sistemas, uso de URIs para la localización de conceptos y formatos de descripciones en otros estándares tecnológicos que sugieren interoperabilidad. Investigaciones futuras pueden explorar mejor el SKOS en ambientes de LD, orientados principalmente al Library Linked Data (LLD) en el ámbito brasileño y teniendo en cuenta iniciativas internacionales más avanzadas.

Palabras clave: Datos vinculados. Datos vinculados a bibliotecas. Sistemas de Organización del Conocimiento. Simple Knowledge Organization System.

1 Introdução

Desde a sua concepção, a World Wide Web passou por estágios evolutivos do escopo que a constitui; dos aspectos que envolvem sua infraestrutura em nível sintático baseada em linguagem HTML original até a amálgama de ordem semântica proposta posteriormente.

A atual denominação web semântica, orientada por diretrizes, padrões, modelos, sistemas conceituais e tecnologias de informação e comunicação, busca elevar o grau semântico em sistemas informacionais a fim de otimizar a capacidade dos computadores em processar e recuperar informações. Em meados dos anos 90, Benjamins inaugura a concepção do World Wide Web Consortium (W3C), cuja finalidade tem sido estabelecer uma comunidade internacional com várias partes interessadas, onde as organizações membros e equipes em tempo integral e o público trabalham juntos para desenvolver padrões abertos da web no sentido de orientar o conjunto tecnológico e as melhores práticas que agregam valor semântico ao ambiente web (World Wide Web Consortium, 2023, tradução nossa).

Dentre as propostas para a web semântica surge em 2006 o *Linked Data* (LD). Segundo Berners-Lee, “A Web Semântica não trata apenas de colocar dados na web. Trata-se de fazer links, para que uma pessoa ou máquina possa explorar a web de dados. Com os *linked data*, quando você tiver alguns deles, poderá encontrar outros dados relacionados” (2006, tradução nossa). Isso corresponde ao conjunto de boas práticas por recomendações com vista a ligação semântica entre dados estruturados (no ambiente web). Os projetos em LD promovem a navegação, exploração e descoberta de recursos informacionais conectados entre si pela associação harmônica de padrões tecnológicos interoperáveis.

Embora as iniciativas de LD retratem um cenário otimista em relação ao avanço da camada semântica da web, é necessário discutir constantemente a temática em seus aspectos teóricos e práticos, bem como entender quais são as tecnologias que agregam semântica.

No contexto da Ciência da Informação e áreas afins, a Biblioteconomia tem contribuído com ambientes de LD a partir de práticas antigas de Organização do Conhecimento atuando no controle terminológico e conceitual, base para o avanço de criações semânticas em sistemas de informação (Martínez-Ávila; San Segundo; Zurian, 2014; Martínez-Ávila, 2015).

Um exemplo é a aplicação dos Sistemas de Organização do Conhecimento (SOC) que incluem: sistemas de classificação, códigos de classificação, listas de cabeçalhos ou cabeçalhos de assuntos, listas de descritores, listas de autoridades, anéis de sinônimos, mapas conceituais, mapas de tópicos, diretórios de buscadores, *folksonomias*, taxonomias, tesouros e ontologias (Hodge, 2000; Mazzocchi, 2018).

Isto posto, o modelo Simple Knowledge Organization System (SKOS), criado em 2009, possui a finalidade de simplificar um Sistema de Organização do Conhecimento em estruturas por conceitos compostas por: URIs que rotulam cadeias lexicais em um ou mais idiomas e notações atribuídas; por códigos lexicais documentadas com vários tipos de notas e que vinculam outros conceitos organizados em hierarquias informais e redes de associação agregando esquemas de conceito; e por agrupamento de coleções rotuladas e/ou ordenadas e mapeadas para conceitos em outros esquemas (Miles; Bechhofer, 2009, tradução nossa). Esse tipo de representação tem por finalidade a elaboração de vocabulários estruturados em conceitos legíveis por sistemas de informação a fim de aumentar suas cadeias semânticas no âmbito da recuperação.

Considerando as partes que constituem o LD, Pastor-Sánchez, Martínez-Méndez e Rodríguez-Muñoz (2012) apresentam o modelo de dados SKOS como uma ontologia de resultados satisfatórios e amplamente aplicada no ambiente da web semântica devido a sua capacidade de fornecer um modelo para representar a estrutura básica e o conteúdo dos SOC em esquemas conceituais legíveis por computador.

Riley (2017) destaca que, embora o modelo SKOS seja útil principalmente no caso de uso restrito à sua finalidade de representar um Sistema de Organização do Conhecimento, várias propriedades podem ser usadas para conectar conceitos que se relacionam entre si e são amplamente usados em outras aplicações do LD.

Desse modo, surge o problema de pesquisa: como utilizar o modelo SKOS para representar conceitos de Sistemas de Organização do Conhecimento e como essas representações conceituais integram iniciativas de LD?

O artigo busca discutir o modelo SKOS como tecnologia que integra iniciativas de LD a partir da representação de três conceitos e suas relações. Com base nos trabalhos de Summers *et al.* (2008), Miles e Bechhofer (2009), Santos, Cervantes e Fujita (2014), Pastor-Sánchez, Martínez-Méndez e Rodríguez-Muñoz (2012), Riley (2017), Ramalho (2017) e alguns documentos do consórcio W3C e da LC criados para orientar representações em SKOS e estabelecer as boas práticas para o LD, propomos a representação em SKOS de conceitos exatos e similares em SOC distintos. Desse modo, iremos discutir alguns aspectos que orientam as possibilidades de ligações técnicas entre os dados de SOC distintos por meio do modelo SKOS.

Para Barbosa e Viera (2022) as estruturas conceituais representadas em SKOS se estabelecem a partir de duas perspectivas: intra-vocabulários e inter-vocabulários. A primeira se refere as relações semânticas entre os termos de um mesmo vocabulário (representações de termos e variações linguísticas, relacionamentos entre grupos e subgrupos de conceitos); a segunda está associada ao mapeamento entre vocabulários distintos que, conforme Barbosa e Viera (2022), se torna um indicativo de interoperabilidade devido os vínculos estabelecidos entre conceitos de diferentes vocabulários. Nesse contexto, entende-se que o SKOS possui capacidade para vincular dados entre vocabulários distintos (inter-vocabulários) de Sistemas de Organização do Conhecimento.

Como resultados, busca avançar nas discussões teóricas e práticas que envolvem representações em SKOS no intuito de estabelecer ligações conceituais inter-vocabulários como parte integrativa do LD.

2 Referencial teórico

Em seu artigo sobre Linked Data, Marcondes (2012) introduz um exemplo hipotético onde o protagonista da sua história inicia uma busca por determinado assunto. Devido a curiosidade em saber mais sobre a temática pesquisada, ele começa a ser orientado a navegar por links conectados onde todos os assuntos estão relacionados entre si. Em âmbito geral, a concepção do Linked Data carrega essa ideia: um conjunto de dados estruturados por uma teia de hiperlinks que estabelecem ligações semânticas com outros dados relacionados.

Benjers-Lee (2006, tradução nossa) apresenta quatro “regras” que orientam a construção de um ambiente em Linked Data. Na verdade, o autor se refere a essas regras como **expectativas de comportamento** que precisam ser consideradas, pois não segui-las também indicaria a possível inviabilidade de tornar os dados interconectados, limitando o reuso e afetando o valor semântico acrescentado pela web.

As expectativas de comportamento referenciadas correspondem às boas práticas para o Linked Data: a) a primeira diz respeito a identificação de coisas por Uniform Resource Identifier (URI) - Identificador Uniforme de Recurso composto por um conjunto universal de símbolos que identificam um recurso na web; b) a segunda orienta o uso do protocolo de comunicação Hypertext Transfer Protocol (HTTP) e uso de URIs - como por exemplo: Life Science Identifier (LSID), Handle (identificador persistente de objetos digitais e outros recursos na web), eXtensible Resource Identifier (XRI) e o Digital Object Identifier (DOI) - a fim de que as pessoas possam procurar nomes únicos; c) a terceira indica o uso de ontologias/padrões tecnológicos como o Resource Description Framework (RDF), RDF Schema, Ontology Web Language (OWL) para incluir relacionamentos entre os termos sempre que alguém buscar por um URI e; d) a quarta indicação sugere que se estabeleçam links para outros URIs (em sites de hipertexto, geralmente é considerado falta de etiqueta não criar links para material externo relacionado, pois o valor de suas próprias informações depende em grande parte do link a que elas estão vinculadas, bem como do valor inerente das informações contidas na página da web (Benjers-Lee, 2006, tradução nossa). Assim, tem-se o formato de um ambiente desejável em Linked Data, entendido como uma expansão ao conceito da web semântica.

A ideia de Linked Data tem sua concepção de aplicação no ambiente web. No contexto das bibliotecas, o conceito de Library Linked Data (LLD), aqui referenciado como qualquer tipo de dados de biblioteca que descreva recursos por meio de padrões, vem sendo constituído a partir de relatórios de trabalho a fim de orientar agentes e atores de bibliotecas à aplicação de possíveis práticas de LD em bibliotecas. O Library Linked Data Incubator Group Final Report (Baker et al., 2011, tradução nossa) propôs um documento que orienta a concepção do LLD. São elas:

- Que os líderes das bibliotecas identifiquem conjuntos de dados como possíveis candidatos para exposição antecipada como Linked Data e promovam uma discussão sobre dados abertos e direitos;
- Que os órgãos de padronização de bibliotecas aumentem a participação da biblioteca na padronização da web semântica, desenvolvam padrões de dados de biblioteca que sejam compatíveis com Linked Data e disseminem os padrões de design de melhores práticas adaptados aos Linked Data de Biblioteca;

- Que os designers de dados e sistemas projetem serviços de usuário aprimorados com base nos recursos de Linked Data, criem URIs para os itens em conjuntos de dados de biblioteca, desenvolvam políticas para gerenciar RDF e seus URIs e expressem dados de biblioteca reutilizando ou mapeando elementos de Linked Data existentes e;
- Que os bibliotecários e arquivistas preservem os conjuntos de elementos Linked Data e valorizem sua estrutura aplicando a experiência da biblioteca em curadoria e preservação de longo prazo para conjuntos de dados em Linked Data.

Em se tratando dos processos técnicos em biblioteca, a descrição de dados oriunda da prática da catalogação nasceu, segundo Alves e Santos (2013), da necessidade em se estabelecer regras para a construção de catálogos. Mas, além disso, foi sendo desenvolvida e aprimorada não apenas como técnica para a construção de catálogos, e sim como disciplina científica com teorias e métodos próprios para o processamento e tratamento descritivo da informação (Alves; Santos, 2013). É preciso levar em consideração que, dentro da conjuntura da Organização do Conhecimento em geral, existe uma série de modelos conceituais, códigos e padrões que dão o seu caráter científico.

Dos catálogos manuais, até a mudança de paradigma oriunda dos catálogos automatizados, a prática da catalogação tem discutido novos meios e estruturas para acompanhar os avanços introduzidos desde a concepção da web semântica. A título de exemplo, as discussões de novas propostas que substituam padrões limitados para bibliotecas são iminentes.

De acordo com Park e Kipp (2019, tradução nossa), o formato MARC - padrão que fornece o mecanismo pelo qual os computadores trocam, usam e interpretam informações bibliográficas - cumpre de maneira satisfatória com o seu objetivo inicial, uma vez que inclui uma descrição rica em metadados a partir da sua estrutura em campos fixos e identificadores de dados importantes sobre um item (visíveis e legíveis por máquinas). Todavia, o MARC não conseguiu avançar no contexto da web semântica devido a sua estrutura rígida e por não ter acompanhado as discussões que estabelecem semântica e ligação de dados sob as novas formas de interação com outros recursos de informação, criando ambientes de dados pouco interativos e cuja finalidade compreende, basicamente, o compartilhamento de dados bibliográficos entre bibliotecas. Uma proposta de substituição do MARC para o entorno bibliotecário compatível com as tecnologias de LD e a web semântica tem sido o BIBFRAME, um modelo que fornece uma base para o futuro da descrição bibliográfica baseada em técnicas de Linked Data. Este padrão, por sua vez, não tem tido sucesso global devido sobretudo as resistências e dinâmicas de trabalho de bibliotecas individuais baseadas em uma longa tradição de MARC (Martínez-Ávila, 2023). Para Serra e Santarém Segundo (2019), ao copiar ou produzir registros para os seus catálogos baseados em MARC, as bibliotecas estão criando ambientes em silos de dados, onde

os metadados não possuem conexão com os demais dados presentes no catálogo, com catálogos de outras bibliotecas ou com a web.

Uma alternativa discutida na literatura para solucionar o problema dos catálogos versam no contexto da abertura dos dados e pela inclusão de outras tecnologias que proporcionem a ligação desses dados. Pastor-Sánchez (2012, tradução nossa) ressalta que o desenvolvimento de modelos de metadados e sua implantação por meio de padrões da web semântica abrem novas perspectivas para o uso de Sistemas de Organização do Conhecimento em geral. Desse modo, tesouros, arquivos de autoridade ou listas de cabeçalhos de assunto são mais valiosos se forem publicados sob o paradigma dos dados abertos vinculados, pois facilita a interoperabilidade e, conseqüentemente, o reuso e aplicação em qualquer outro campo (Pastor-Sánchez, 2012, tradução nossa). E completa:

dentro da diversidade de vocabulários controlados, os arquivos de controle de autoridade são talvez os que mais se vinculam ao processo de catalogação, pois sua gênese ocorre nesse processo. Embora a sua criação e estruturação não se baseiem nos princípios que regem a organização do conhecimento, é indiscutível a sua incardinação nesta disciplina. As classificações, tesouros e cabeçalhos de assunto pressupõem uma estrutura conceitual interna que, embora por vezes tenha sido utilizado como corpus documental para a sua elaboração, precede a indexação. As autoridades são identificadas, extraídas e normalizadas durante a descrição bibliográfica, cujo controle é centralizado em um arquivo ou registro de controle de autoridades internas ou externas” (Pastor-Sánchez, 2012, tradução nossa).

Em bibliotecas e organizações a nível internacional, é possível observar alguns exemplos consistentes em relação a modelos, padrões e outras tecnologias que envolvem LD. Na perspectiva da publicação de dados de autoridade como LD, Assumpção (2018) compilou algumas iniciativas em um cenário global, e que passa, obrigatoriamente, pelas etapas: definição de identificador único de recurso/objeto, representação dos dados utilizando um modelo/padrão comum, estabelecimento de links para outros conjuntos de dados e publicação dos dados.

Na complexidade dos padrões que buscam integrar o LD, o SKOS se apresenta como um modelo de dados que compartilha e vincula um Sistema de Organização do Conhecimento por esquemas conceituais de sua estrutura em um vocabulário de dados. Pastor-Sánchez, Martínez-Méndez e Rodríguez-Muñoz (2012, tradução nossa) entendem que é necessário determinar com mais precisão o verdadeiro potencial do SKOS, avaliando seu impacto, evolução e aplicação em conjuntos de dados associados a linguagens documentais de todos os tipos para que, dessa forma, o SKOS possa ser avaliado como parte do ecossistema da web semântica ou apenas como fenômeno passageiro. Todavia, os autores acreditam que o modelo SKOS pode ser visto como

uma ontologia satisfatória e amplamente aplicada no ambiente da web semântica, pois fornece um modelo para representar a estrutura básica e o conteúdo de esquemas conceituais, tais como listas de cabeçalhos de assuntos, taxonomias, esquemas de classificação, tesouros e qualquer outro tipo de Sistema de Organização do Conhecimento.

Dentre os componentes dos elementos que compõem o SKOS estão os conceitos e classes, as relações semânticas e as propriedades dos conceitos. No artigo, limita apresentar tecnicamente (tabelas 1 e 2) apenas elementos do modelo SKOS aplicados aos conceitos a serem representados em SKOS. A seguir:

Tabela 1 – Conceito e Classes no modelo de dados SKOS

skos:Concept
URI: http://www.w3.org/2004/02/skos/core#Concept
Descrição: conceito
skos:Collection
URI: http://www.w3.org/2004/02/skos/core#Collection
Descrição: coleções de conceitos
skos:ConceptScheme
URI: http://www.w3.org/2004/02/skos/core#ConceptScheme
Descrição: esquema conceitual

Fonte: Adaptado pelos autores a partir de Miles e Bechhofer (2009) Isaac e Summers (2009)

Na Tabela 1 estão dispostos os conceitos e as classes de conceitos no SKOS. Estes são representados pelos elementos: `skos:Concept`, `skos:Collection` e `skos:ConceptScheme`. Respectivamente é possível dizer que o primeiro se trata da descrição de um conceito e suas características em um SOC; o segundo descreve uma coleção de conceitos; e o terceiro envolve a descrição de um esquema conceitual. Na Tabela 1, todos os elementos apresentam um identificador único de localização que remete ao documento SKOS Simple Knowledge Organization System Namespace Document - HTML Variant (Miles, Bechhofer, 2009, tradução nossa). Isso também se repete na Tabela 2 quando trata dos elementos de propriedades do modelo. A seguir:

Tabela 2 – Propriedades no modelo de dados SKOS

<i>skos:prefLabel</i>
URI: http://www.w3.org/2004/02/skos/core#prefLabel
Tipo: Rótulo lexical
Descrição: rótulo preferido
<i>skos:altLabel</i>
URI: http://www.w3.org/2004/02/skos/core#altLabel
Tipo: Rótulo lexical
Descrição: rótulo alternativo
<i>skos:broader</i>
URI: http://www.w3.org/2004/02/skos/core#broader
Tipo: Relação semântica
Descrição: rótulo mais amplo
Inverso de: skos:narrower
<i>skos:narrower</i>
URI: http://www.w3.org/2004/02/skos/core#narrower
Tipo: Relação semântica
Descrição: rótulo mais estreito
Inverso de: skos:broader
<i>skos:notation</i>
URI: http://www.w3.org/2004/02/skos/core#notation
Descrição: notação
<i>skos:closeMatch</i>
URI: http://www.w3.org/2004/02/skos/core#closeMatch
Tipo: Propriedade de mapeamento
Descrição: tem correspondência próxima
<i>skos:exactMatch</i>
URI: http://www.w3.org/2004/02/skos/core#exactMatch
Tipo: Propriedade de mapeamento
Descrição: tem correspondência exata
<i>skos:inScheme</i>
URI: http://www.w3.org/2004/02/skos/core#inScheme
Tipo: Esquema de um conceito
Descrição: está em esquema
Faixa: skos:ConceptScheme
<i>skos:topConceptOf</i>
URI: http://www.w3.org/2004/02/skos/core#topConceptOf
Tipo: Esquema de um conceito
Descrição: é o conceito principal no esquema
Super propriedade: skos:inScheme

Fonte: Adaptado pelos autores a partir de Miles e Bechhofer (2009) Isaac e Summers (2009)

Para Riley (2017, tradução nossa), o SKOS possui várias propriedades que podem ser usadas para conectar conceitos que se relacionam entre si e são amplamente usados em outras aplicações do LD, como por exemplo, as propriedades frequentemente usadas e destinadas para esse fim; skos:broadMatch, skos:closeMatch e skos:exactMatch.

Em resumo, o modelo SKOS possui a capacidade de codificar relações de seus elementos por meio de classes e propriedades, onde toda sua estrutura está baseada no conceito. Este pode ser descrito como um único conceito, classe, categoria, coleção de conceitos ou mesmo

um esquema conceitual. Ao que trata suas propriedades, adquirem marcações para rótulos de conceitos e classes, relacionamentos, notas descritivas e mapeamentos.

3 Percurso metodológico

Sobre os aspectos metodológicos, a pesquisa tem natureza aplicada, uma vez que gera uma base de conhecimentos direcionados aos aspectos sociais (Boaventura, 2004). Tem caráter teórico, pois se constrói à luz de outras pesquisas para discutir as temáticas que envolvem os avanços do LD e do modelo conceitual SKOS. Sobre os objetivos, a pesquisa é exploratória e descritiva: a primeira por se tratar da formulação de problemas orientados ao melhor entendimento dos temas estudados e a segunda pela descrição das características dos conceitos indicados e representados no modelo conceitual SKOS. Tem abordagem qualitativa, pois a partir da descrição dos fenômenos estudados gera um produto de visão subjetiva e interpretação dos resultados pela percepção dos fenômenos estudados (Triviños, 2007).

Para o estudo, foram eleitos os seguintes conceitos: **“tecnologias da informação e comunicação” do Tesouro Brasileiro de Ciência da Informação** com base no trabalho de Ramalho (2017), **“Tecnologia da informação” do Catálogo de Autoridades de Assunto da Biblioteca Nacional (Brasil)** e **“Information technology” da Library of Congress Subject Headings**. A partir desses conceitos eleitos, foram discutidas suas relações e como esse vínculo (que também denota a concepção de LD) se estabelece por meio dos elementos do modelo SKOS.

Para as análises foram utilizados os seguintes instrumentos: o programa XML Copy Editor para criação de uma base sintática utilizando linguagem de marcação e os elementos do modelo SKOS que estabelecem ligações entre conceitos e outros recursos de padrões tecnológicos; o W3C RDF Validator que valida operações descritas no formato SKOS e o programa Diagrams.net para criar as representações gráficas em SKOS.

4 Discussões e resultados

Para entender a forma como esses conceitos foram apresentados em seus sistemas, elaboramos as Tabelas 3, 4 e 5 (em sequência) para melhor visualização e discussão:

Tabela 3 – “tecnologias da informação e comunicação” do Tesouro Brasileiro de Ciência da Informação

tecnologias da informação e comunicação

ING: informationtechnology
ESP: tecnología de la información, tecnología computacional

Usado para: (UP)
tecnologias da informação
TICs

Termo Relacionado (TR)
normas e protocolos
tecnologias assistivas
telecomunicações
ciência da informação
gestão nas tics
organização do conhecimento
recuperação da informação

Nota de Escopo (NE):
Abarca todas as formas de tecnologia usadas para criar, armazenar, intercambiar e usar a informação em suas várias formas - ASIST, p. 69.

Termo Geral (TG)
CAT: 5 Tecnologias da Informação e Comunicação - TICs

Fonte: Elaborado pelos autores a partir de Pinheiro e Ferrez (2014)

A Tabela 3 representa o recorte do conceito “tecnologias da informação e comunicação” do Tesouro Brasileiro de Ciência da Informação, baseado no trabalho de Ramalho (2017). O Tesouro Brasileiro de Ciência da Informação desenvolvido originalmente por Pinheiro e Ferrez e disponibilizado no site do IBICT disponível em: <http://sitehistorico.ibict.br/publicacoes-e-institucionais/tesouro-brasileiro-de-ciencia-da-informacao-1>, foi convertido por Santos, Cervantes e Fujita (2014) e disponível em: <http://www.uel.br/revistas/informacao/tbci/vocab/index.php>. Em seguida, temos a Tabela 4 que corresponde ao conceito “Tecnologias da informação” do Catálogo de Autoridades de Assunto da Biblioteca Nacional (2023). Trata-se de um cabeçalho de assuntos que compõe o SOC.

Tabela 4 – “Tecnologias da informação” do Catálogo de Autoridades de Assunto da Biblioteca Nacional

Tecnologia da informação (Subdividido geograficamente)

Usado para (UP):
Technology, Information
IT (Information technology)

Termo geral (TG):
Tecnologia
Telemática

Termo relacionado (TR):
Gestão do conhecimento
Superestrada da informação

Nota geral pública (NE):
Usado para obras que tratam da aquisição, processamento, armazenamento e disseminação de informações vocais, pictóricas, textuais e numéricas através da microeletrônica, computadores e telecomunicação

Fonte: Elaborado pelos autores a partir da Biblioteca Nacional – Brasil (2023)

Em seguida temos a Tabela 5 que representa o conceito “Information technology” na Library of Congress Subject Headings (2023). A seguir:

Tabela 5 – “Information technology” da Library of Congress Subject Headings

Information technology (May Subd Geog)

Here are entered works on the acquisition, processing, storage and dissemination of vocal, pictorial, textual and numeric information by microelectronics, computers and telecommunication.

Used For (UF):
IT (Information technology)

Brother Term (BT):
Technology
Telematics

Related Term (RT):
Information superhighway
Knowledge management

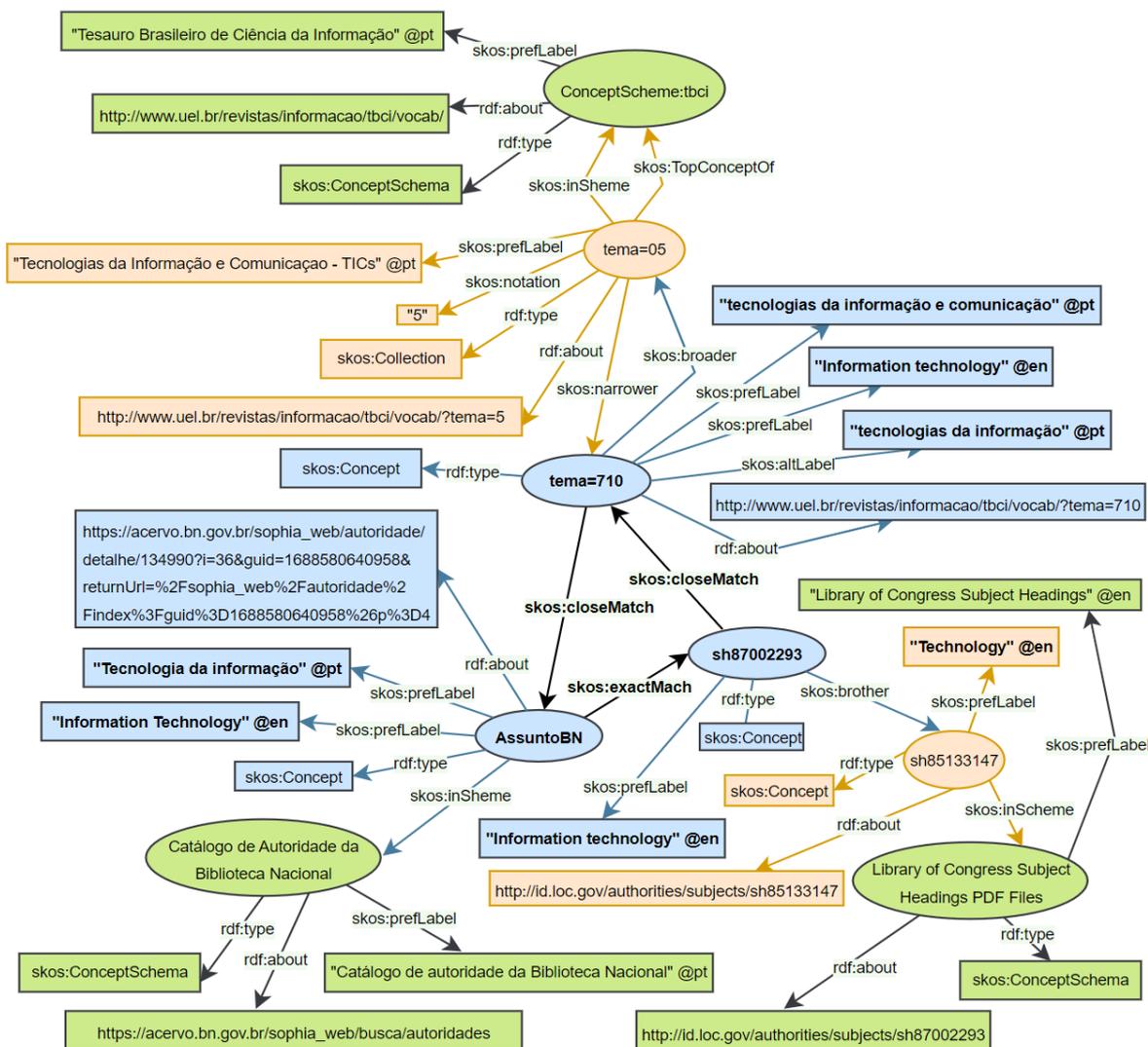
Escope Note (AS):
Subdivision Information technology under names of individual corporate bodies, and under disciplines, types of industries, services, and organizations, e.g. Construction industry--Information technology

NarrowerTerm (NT):
Cyberinfrastructure

Fonte: Elaborado pelos autores a partir da Library of Congress Subject Headings (2023)

Considerando as Tabelas 3, 4 e 5, apresenta-se a Figura 1 que diz respeito a figura correspondente às representações dos conceitos sob a forma do SKOS:

Figura 1 – Representação em SKOS dos conceitos “tecnologias da informação e comunicação” do Tesouro Brasileiro de Ciência da Informação, “Tecnologias da informação” do catálogo de autoridade de assuntos da Biblioteca Nacional e “Information technology” dos Library of Congress SubjectHeadings e seus relacionamentos



Fonte: Adaptação feita pelos autores a partir de Ramalho (2017); elaborado pelos autores (2023)

A Figura 1 apresenta a ligação dos conceitos por meio do mapeamento a partir de duas propriedades: `skos:closeMatch` e `skos:exactMatch`. De acordo com Miles e Bechhofer (2009, tradução nossa), essas propriedades são usadas para vincular conceitos SKOS em diferentes esquemas: o `skos:closeMatch` indica que dois conceitos são suficientemente semelhantes para que possam ser usados de forma intercambiável em alguns aplicativos de recuperação de informações e o `skos:exactMatch` que indica um alto grau de confiança de dois conceitos que

podem ser usados de forma intercambiável em uma ampla gama de aplicativos de recuperação de informações.

Na Figura 2, podemos visualizar essa representação a partir do software XML Copy Editor, importante para a validação em RDF das sentenças em SKOS.

Figura 2 - Representação em SKOS dos conceitos “tecnologias da informação e comunicação” do “Tesouro Brasileiro de Ciência da Informação”, “Tecnologias da Informação” do “Catálogo de Autoridades da Biblioteca Nacional” e “Information Technology” dos “Library of Congress SubjectHeadings” expressos em XML/RDF.

```
1 <rdf:RDF
2
3 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4 xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
5 xmlns:skos="http://www.w3.org/2004/02/skos/core#"
6 xmlns:map="http://www.w3c.rl.ac.uk/2003/11/21-skos-mapping#"
7 xmlns:dct="http://purl.org/dc/terms/"
8 xmlns:dc="http://purl.org/dc/elements/1.1/"
9
10 <skos:Concept rdf:about="http://www.uel.br/revistas/informacao/tbci/vocab/?tema=710">
11 <skos:prefLabel xml:lang="pt">tecnologias da Informação e comunicação</skos:prefLabel>
12 <skos:prefLabel xml:lang="en">Information technology</skos:prefLabel>
13 <skos:altLabel xml:lang="pt">tecnologias da Informação</skos:altLabel>
14 <skos:inScheme rdf:resource="http://www.uel.br/revistas/informacao/tbci/vocab/">
15 <skos:broader rdf:resource="http://www.uel.br/revistas/informacao/tbci/vocab/?tema=5">
16 <skos:closeMatch xml:lang="pt">Tecnologia da Informação</skos:closeMatch>
17 </skos:Concept>
18
19 <skos:Concept rdf:about="https://acervo.bn.gov.br/sophia_web/busca/autoridades">
20 <skos:prefLabel xml:lang="pt">Tecnologia da Informação</skos:prefLabel>
21 <skos:prefLabel xml:lang="en">Information technology</skos:prefLabel>
22 <skos:inScheme rdf:resource="https://acervo.bn.gov.br/sophia_web/busca/autoridades">
23 <skos:broader rdf:resource="https://acervo.bn.gov.br/sophia_web/busca/autoridades">
24 <skos:exactMatch xml:lang="en">Information Technology</skos:exactMatch>
25 </skos:Concept>
26
27 <skos:Concept rdf:about="http://id.loc.gov/authorities/subjects/sh87002293">
28 <skos:prefLabel xml:lang="pt">Information Technology</skos:prefLabel>
29 <skos:inScheme rdf:resource="http://id.loc.gov/authorities/subjects">
30 <skos:broader rdf:resource="http://id.loc.gov/authorities/subjects/sh85133147">
31 <skos:closeMatch xml:lang="pt">Tecnologias da Informação</skos:closeMatch>
32 </skos:Concept>
33
34 </rdf:RDF>
```

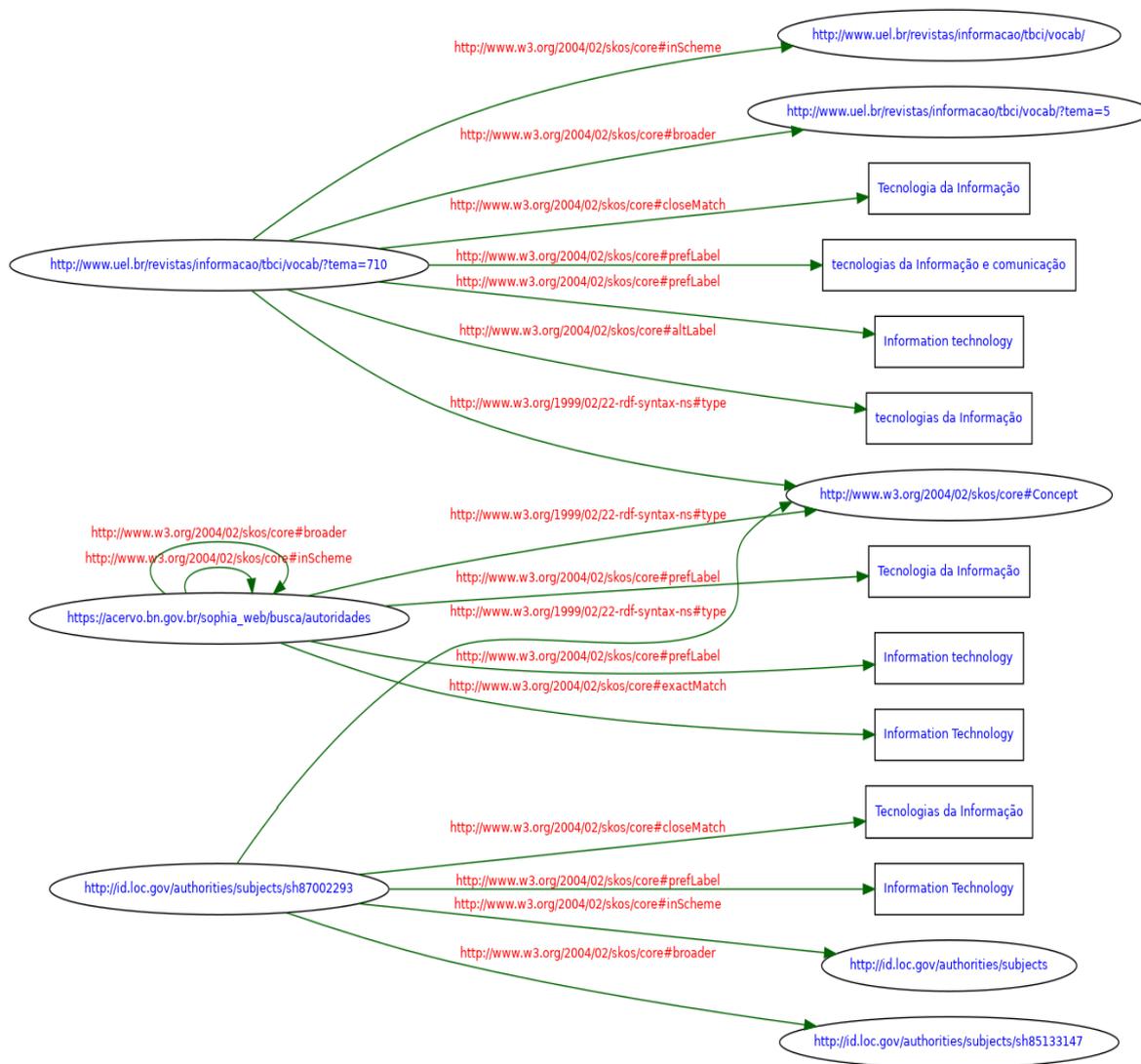
Fonte: Elaborado pelos autores a partir dos metadados do TBCI Eletrônico (Santos, Cervantes, Fujita, 2014)

Por meio da estrutura disponível na Figura 2, foi possível validar em RDF as sentenças dos relacionamentos entre os três conceitos que constituíram o escopo da pesquisa, conforme apresenta a Figura 3.

Figura 3 – Gráfico dos conceitos “tecnologias da informação e comunicação”, “Tecnologia da Informação” e “Information Technology” em SKOS

Resultados da validação

Seu documento RDF foi validado com sucesso.

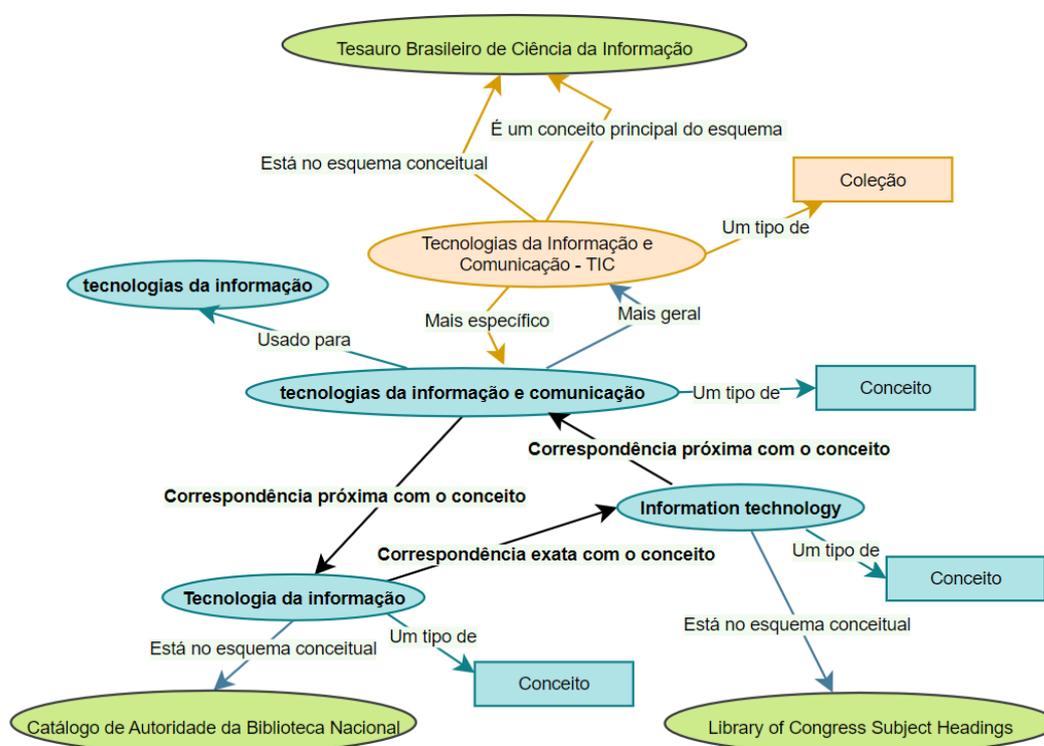


Fonte: Validador RDF (World Wide Web Consortium, 2023)

De modo a aproximar a descrição da Figura 3 a uma linguagem humana, temos: na primeira elipse (à esquerda), uma URL que é um conceito e faz parte do esquema conceitual “Tesouro Brasileiro de Ciência da Informação” tem como termo preferido em português o conceito “tecnologias da informação e comunicação” e termo preferido para o conceito em inglês “Information technology”, termo relacionado ao conceito “Tecnologia da informação” e possui uma relação de proximidade com o conceito “Tecnologias da informação”. Este, por sua vez, na segunda elipse (à esquerda) está descrito por uma URL, é um conceito que faz parte do esquema conceitual do “Catálogo de Autoridade da Biblioteca Nacional” tem como termo

preferido em português o conceito “Tecnologias da informação” e termo preferido para o conceito em inglês “Information technology” e que, por sua vez, possui uma relação de exatidão com o conceito “Information technology”. Esse último, indicado na terceira elipse (à esquerda) retirado dos “Library of Congress Subject Headings”, está descrito por uma URI e corresponde a um esquema conceitual descrito por uma nova URI, seu termo preferido está expresso pelo conceito “Information technology” e este possui uma relação de proximidade com o conceito “tecnologias da informação e comunicação” do TBCI. Nesse sentido, a Figura 4 torna um pouco mais compreensível esta visualização em linguagem natural.

Figura 4 – Representação gráfica do SKOS descrita em linguagem natural



Fonte: Elaborado pelos autores (2023).

5 Considerações finais

A partir do estudo foi possível apresentar o modelo SKOS como tecnologia que integra iniciativas de Linked Data (LD) a partir da representação de três conceitos e suas relações. Dessa forma, entendemos que o objetivo proposto foi alcançado levando em consideração os pontos a seguir.

O conceito “tecnologias de informação e comunicação” possui em sua estrutura características aplicáveis ao modelo SKOS. Entretanto, notamos a ausência de URIs para a

localização de dados. Mesmo havendo a existência do endereço que corresponde ao conceito e suas relações, estes são apresentados sob a forma de uma URL (não recomendável para LD) ao invés de uma URI (recomendação apropriada). A criação de identificadores únicos para objetos envolve questões que não pretendemos discutir neste artigo. Entretanto, a ausência de URIs não impossibilitou a realizações de ligações com outras URIs exatas. De todo modo, essa indicação é importante, pois traria maior potência para representações desse tipo.

O mesmo se apresenta no conceito “Tecnologias da informação”, do catálogo de autoridade de assunto da Biblioteca Nacional - Brasil. Aqui temos uma situação mais complicada. Não há, no catálogo da Biblioteca Nacional, nenhuma apresentação de práticas que envolvem LD. O termo se apresenta no catálogo de maneira apenas descritiva. Isso pode ser justificado por sua própria constituição e finalidade sendo este um cabeçalho de assunto. Não há indicação de dados abertos em linguagem computacional própria para o ambiente que operem em SKOS (ou outras tecnologias para o LD), nem indicações de URIs para os assuntos. Ao que parece, a própria constituição do cabeçalho de assuntos da Biblioteca Nacional está representada a partir dos Library of Congress Subject Headings.

Por sua vez, todos os assuntos dos Library of Congress Subject Headings possuem indicação de URIs para todos os seus assuntos, além de infraestrutura em SKOS e outras descrições em padrões tecnológicos que sugerem certo grau de interoperabilidade. A LC propõe ainda o uso do MADS/RDF (Library of Congress, 2015) que foi projetado para complementar o SKOS e, como tal, é formalmente mapeado para o SKOS/RDF oferecendo, principalmente, suporte a dados de autoridade usados e necessários na comunidade da Library and Information Science (CI) e seus sistemas de tecnologia.

No contexto da Ciência da Informação, o modelo SKOS é considerado simples em sua estrutura, uma vez que corresponde à representação de um Sistema de Organização do Conhecimento. Analisando seus conceitos, propriedades e tipos de relacionamento, é possível visualizar a mesma lógica que está por trás da organização de um SOC. Entretanto, quando se trata de pensá-lo como uma tecnologia aplicada que agregue valor ao contexto do LD, percebe-se que existem grandes desafios e limitações. Podemos listar:

- a temática ainda é insuficientemente explorada na Ciência da Informação brasileira. Podemos citar como exemplo a dificuldade em encontrar suficientes pesquisas de autores brasileiros que discutem a temática do SKOS na Ciência da Informação sob os aspectos teóricos, metodológicos e práticos levando em consideração alguns avanços de iniciais internacionais;
- necessidade de conhecimentos técnicos avançados: mesmo sendo um modelo simples, exige o conhecimento de linguagens de marcação e uma série de tecnologias computacionais que, por conseguinte, darão condições para a interoperabilidade com outros sistemas ao ponto de criar redes com outras tecnologias de LD. Isso pode ser uma dificuldade no contexto bibliotecário;

- depende do engajamento interdisciplinar para avançar em suas teorias e constituir propostas metodológicas que demonstrem o verdadeiro potencial prático do modelo SKOS. Muitas vezes depende da vontade de informáticos em trabalhar sobre temáticas da CI.

Assim, podemos sugerir em pesquisas futuras, estudos que explorem os aspectos de melhoria aos pontos levantados a fim de que os SOC apresentados melhorem suas estruturas na ligação de dados. Além disso, entendemos que os assuntos que correspondem o conjunto de práticas para ambientes de LD, orientados principalmente ao conceito específico do Library Linked Data, devem ser discutidos e estudados (em teoria e prática) no âmbito brasileiro, levando em consideração as iniciativas internacionais como a LCSH. Essas já em estágios mais avançados.

Referências

ALVES, R. C. V.; SANTOS, P. L. V. A. da C. **Metadados no domínio bibliográfico**. Rio de Janeiro: Intertexto, 2023.

ARAKAKI, F. A. **Linked Data: ligação de dados bibliográficos**. 2016. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Filosofia e Ciências, Marília, 2016. Disponível em: <https://repositorio.unesp.br/handle/11449/147979>. Acesso em: 09 jul. 2023.

ASSUMPÇÃO, F. S. **Modelo para a publicação de dados de autoridade como Linked Data**. 2018. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Filosofia e Ciência, Marília, 2018. Disponível em: <http://hdl.handle.net/11449/152759>. Acesso em: 10 jul. 2023.

BAKER. *et al.* **Library Linked Data Incubator Group: Final Report**. W3C Incubator Group Report 25 October 2011. Cambridge: W3C, 2011. Disponível em: <https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>Acesso em: 10 jul. 2023.

BARBOSA, E. R.; VIERA, A. F. G. Relações semânticas e interoperabilidade em tesouros representados em SKOS. **Informação & Informação**, Londrina, v. 27, n. 1, p. 229-252, jan./mar. 2022. DOI: <https://doi.org/10.5433/1981-8920.2022v27n1p229>

BERNERS-LEE, T. Is your Linked Open Data 5 star? In: BERNERS-LEE, T. **Linked Data**. Cambridge: W3C, 2010. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 08 jul. 2023.

BERNERS-LEE, T. **Linked Data**. Cambridge: W3C, 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>Acesso em: 8 jul. 2023.

BOAVENTURA, E. M. **Metodologia da pesquisa**. São Paulo, SP: Atlas, 2004.

Fundação Biblioteca Nacional. **Catálogo de autoridades**. Rio de Janeiro: FBN, 2023. Disponível em: https://acervo.bn.gov.br/sophia_web/ Acesso em: 10 jul. 2023.

Fundação Biblioteca Nacional. **Relatório de atividades**. Rio de Janeiro: FBN, 2002. Disponível em: <https://www.gov.br/bn/pt-br/aceso-a-informacao-2/auditorias/relatorios-de-gestao/relatorio-de-gestao-2002.pdf/view> Acesso em: 10 jul. 2023.

HODGE, G. **Systems of knowledge organization for digital libraries**: beyond traditional authority files. Washington: The Digital Library Federation; Council on Library and Information Resources, 2000. Disponível em: <https://clir.wordpress.clir.org/wp-content/uploads/sites/6/pub91.pdf> Acesso em: 21 out. 2023.

ISAAC, A.; SUMMERS, E. (ed.). **SKOS Simple Knowledge Organization System Primer**. Cambridge: W3C Working Group Note, 2009. Disponível em: <https://www.w3.org/TR/skos-primer/> Acesso em: 08 jul. 2023.

LIBRARY OF CONGRESS. **Library of Congress Subject Headings PDF Files**. Washington, D.C.: LC, 2023. Disponível em: <https://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html> Acesso em: 10 jul. 2023.

LIBRARY OF CONGRESS. **MADS/RDF Primer**. Washington, D.C.: LC, 2015. Disponível em: <https://www.loc.gov/standards/mads/rdf/> Acesso em: 10 jul. 2023.

MARCONDES, C. H. "Linked Data" - dados interligados - e interoperabilidade entre arquivos, bibliotecas e museus na web. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, v. 17, n. 34, p. 171-192, maio/ago., 2012. DOI: <https://doi.org/10.5007/1518-2924.2012v17n34p171>.

MARTÍNEZ-ÁVILA, D. Knowledge Organization in the Intersection with Information Technologies. **Knowledge Organization**, v. 42, n. 7, p. 486-498, 2015. DOI: 10.5771/0943-7444-2015-7-486

MARTÍNEZ-ÁVILA, D. 2023. La muerte y el regreso de Bibframe. **Anuario ThinkEPI**, v. 17, e17a01. DOI: <https://doi.org/10.3145/thinkepi.2023.e17a41>

MARTÍNEZ-ÁVILA, D.; SAN SEGUNDO, R.; ZURIAN, F. A. Retos y oportunidades en organización del conocimiento en la intersección con las tecnologías de la información. **Revista Española De Documentación Científica**, n. 3, v. 37, p. 1-13, 2014. DOI: <https://doi.org/10.3989/redc.2014.3.1112>

MAZZOCCHI, F. Knowledge organization system. **Knowledge Organization**, v. 45, n. 1, p. 54-78, 2018. Disponível em: <https://www.isko.org/cyclo/kos> Acesso em: 21 dez. 2023

MILES, A.; BECHHOFFER, S (ed.). **SKOS Simple Knowledge Organization System Reference**. Cambridge: W3C, 2009. Disponível em: <https://www.w3.org/TR/2009/REC-skos-reference-20090818/> Acesso em: 08 jul. 2023.

PARK, H.; KIPP, M. Library Linked Data Models: Library Data in the Semantic Web. **Cataloging & Classification Quarterly**, v. 57, n. 5, p. 261-277, 2019. DOI: <https://doi.org/10.1080/01639374.2019.1641171>.

PASTOR SÁNCHEZ, J. A. Ontologías de control de autoridades en el ámbito de los datos abiertos enlazados. **Análisis de Tendencias En Informacion y Documentacion**, v. 7, n. 1, p. 184-

188, 2013. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=4234754>Acesso: 10 jul. 2023.

PASTOR-SANCHEZ, J. A.; MARTINEZ-MENDEZ, F. J.; RODRIGUEZ-MUÑOZ, 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, p. 245-253, 2012. DOI: <https://doi.org/10.3145/epi.2012.may.04>.

PINHEIRO, L. V. R.; FERREZ, H. D. **Tesouro Brasileiro de Ciência da Informação**. Rio de Janeiro; Brasília: Ibict, 2014. Disponível em: <http://sitehistorico.ibict.br/publicacoes-e-institucionais>Acesso em: 09 jul. 2023.

RAMALHO, R. A. S. Representação SKOS da categoria tecnologia da informação e comunicação do tesouro brasileiro de ciência da informação: um estudo preliminar. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v.10, n.2, ago./dez. 2017. Disponível em: <https://revistas.ancib.org/index.php/tpbci/article/view/417>Acesso em: 09 jul. 2023.

RILEY, Jenn. **Understanding Metadata: what is metadata, and what is it for?** 2017. Baltimore: National Information Standards Organization (NISO), 2017. Disponível em: <https://www.niso.org/publications/understanding-metadata-2017>Acesso em: 10 jul. 2023.

SANTOS, J. C.; CERVANTES, B.; FUJITA, M. S. L. (org.). **Tesouro brasileiro de Ciência da Informação: representação online**. Autora do tesouro: coordenação; Lena Vânia Ribeiro Pinheiro e colaboração; Helena Dodd Ferrez. Londrina: UEL, 2014. Disponível em: <http://www.uel.br/revistas/informacao/tbci/vocab/sobre.php>Acesso em: 10 jul. 2023.

SERRA, L. G.; SANTARÉM SEGUNDO, J. E. **Dos silos de dados à web dos dados: bibliotecas e o linked data**. Inf. Inf., Londrina, v. 26, n. 2, p. 625–645, abr./jun. 2021. DOI: <https://doi.org/10.5433/1981-8920.2021v26n2p625>

SUMMERS, E. et al. LCSH, SKOS and Linked Data. **DC-2008--Berlin Proceedings - Papers and Project Reports for DC-2008**, p. 22-26, Sept. 2008. Disponível em: <https://dcpapers.dublincore.org/pubs/issue/view/32>Acesso em: 10 jul. 2023.

TRIVIÑOS, A. N. **Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação**. São Paulo: Atlas, 2007.

WORLD WIDE WEB CONSORTIUM. **About us**. [S.l.], 2023. Disponível em: <https://www.w3.org/about/>Acesso em: 10 jul. 2023.

WORLD WIDE WEB CONSORTIUM. **Semantic Web**. 2013. Disponível em: <https://www.w3.org/standards/>Acesso em: 10 jul. 2023.