



Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação

ISSN: 1518-2924

Programa de Pós-graduação em Ciência da Informação -
Universidade Federal de Santa Catarina

Rautenberg, Sandro; Souza, Lucélia de; Kelniar, João Pedro
UMA PROPOSTA DE TESAURO DA CIÊNCIA DA COMPUTAÇÃO:
ORGANIZAÇÃO DE ELEMENTOS DE CONHECIMENTO COM SKOS

Encontros Bibli: revista eletrônica de biblioteconomia
e ciência da informação, vol. 26, e80386, 2021

Programa de Pós-graduação em Ciência da Informação - Universidade Federal de Santa Catarina

DOI: <https://doi.org/10.5007/1518-2924.2021.e80386>

Disponível em: <https://www.redalyc.org/articulo.oa?id=14768130022>

- Como citar este artigo
- Número completo
- Mais informações do artigo
- Site da revista em redalyc.org

UAEM redalyc.org

Sistema de Informação Científica Redalyc
Rede de Revistas Científicas da América Latina e do Caribe, Espanha e Portugal
Sem fins lucrativos acadêmica projeto, desenvolvido no âmbito da iniciativa
acesso aberto

UMA PROPOSTA DE TESAURO DA CIÊNCIA DA COMPUTAÇÃO: ORGANIZAÇÃO DE ELEMENTOS DE CONHECIMENTO COM SKOS

A proposal for computer science thesaurus: organizing knowledge elements with skos

Sandro Rautenberg

Doutor em Engenharia e Gestão do Conhecimento
Universidade Estadual do Centro-Oeste, Guarapuava, PR,
Brasil
srautenberg@unicentro.br

<https://orcid.org/0000-0002-2375-9365> 

Lucélia de Souza

Doutora em Ciência da Computação
Universidade Estadual do Centro-Oeste, Guarapuava, PR,
Brasil;
lucelia@unicentro.br

<https://orcid.org/0000-0002-9190-7411> 

João Pedro Kelnar

Bacharelado em Ciência da Computação
Universidade Estadual do Centro-Oeste, Guarapuava, PR,
Brasil;
bkelnar2017@gmail.com

<http://orcid.org/0000-0002-1235-088X> 

A lista completa com informações dos autores está no final do artigo 

RESUMO

Objetivo: Considera-se a Organização de Conhecimento como atividade interdisciplinar da Ciência da Informação e da Ciência da Computação que possibilita a representação de elementos de conhecimento em ambientes baseados em Web Semântica. Em face da interdisciplinaridade, o artigo apresenta as escolhas e os esforços despendidos no desenvolvimento de um Tesauro da Ciência da Computação.

Método: como pesquisa aplicada, um processo de desenvolvimento de ontologias é utilizado, ao considerar que ontologias e tesouros são Sistemas de Organização de Conhecimento com atividades correlacionadas.

Resultado: baseado no modelo da Web Semântica *Simple Knowledge Organization System* – SKOS, o tesauro é publicado na Web de Dados e pode ser acessado a partir do *endpoint* <http://lod.unicentro.br/sparql>. Admite-se que o público-alvo do Sistema de Organização de Conhecimento desenvolvido são profissionais, professores, pesquisadores, alunos de graduação/pós-graduação, sendo o tesauro um subsídio à comunicação científica entre os referidos atores.

Conclusões: observa-se que os processos de desenvolvimento de ontologias podem ser empregados para o desenvolvimento de tesouros. Ademais, o uso do SKOS como modelo para o desenvolvimento de tesouros mostrou-se adequado, conforme as premissas de organização e representação de recursos digitais do referido modelo.

PALAVRAS-CHAVE: Tesauro. *Simple Knowledge Organization System*. Ciência da Computação. Sistemas de Organização de Conhecimento.

ABSTRACT

Objectives: We consider the Knowledge Organization as an interdisciplinary activity of Information Science and Computer Science that enables the knowledge representation in Semantic Web environments. In the face of this interdisciplinarity, this paper presents the choices and efforts spent to develop a Computer Science Thesaurus.

Methods: as an applied research, an ontology development process is used, considering that ontologies and thesaurus are Knowledge Organization Systems with correlated activities.

Results: based on a Semantic Web model named *Simple Knowledge Organization System* - SKOS, the thesaurus is published on the Web of Data and can be accessed from the <http://lod.unicentro.br/sparql> endpoint. We admit that the audience of the developed Knowledge Organization System is composed by professionals, teachers, researchers, undergraduate/graduate students, and the use of the thesaurus is useful for scientific communication among these actors.

Conclusions: we observed that the ontology development processes can be also used for thesaurus' development. In addition, the use of SKOS as a model for developing thesaurus shown be feasible, according to the premises of digital resources organization and representation.

KEYWORDS: Thesaurus. Simple Knowledge Organization System. Computer Science. Knowledge Organization Systems.

1 INTRODUÇÃO

Uma das importantes temáticas na Ciência da Informação é a organização de conhecimento. Ela vem ultrapassando as fronteiras da referida ciência pela incorporação de insumos advindos de outras disciplinas (MOREIRA, 2019). Por conseguinte, isso enseja novas propostas de pesquisa interdisciplinares, como a que se propõe neste estudo (SANTOS; MOREIRA, 2018).

No âmbito da Ciência da Informação, a organização de conhecimento visa analisar os conceitos e interrelações dos conceitos (hierárquicas, associativas e de equivalência) de um domínio, formalizando-os mediante os Sistemas de Organização de Conhecimento para representação, registro e comunicação de elementos conceituais (CAMPOS; GOMES, 2006; BRÄSCHER; CAFÉ, 2008, p. 8; MOREIRA, 2019).

Pontualmente, um dos Sistemas de Organização de Conhecimento difundidos na Ciência da Informação são os tesauros. Nas palavras de Ferreira e Maculan (2020), como um Sistema de Organização de Conhecimento, os tesauros objetivam a padronização terminológica, reduzindo a polissemia inerente para auxiliar as atividades de representação, comunicação ou recuperação de elementos de conhecimento.

Vale destacar que, ao considerar os avanços tecnológicos, o desenvolvimento e a utilização de tesauros têm-se tornado proeminente. Principalmente, pelo surgimento do *Simple Knowledge Organization System* (SKOS) como um modelo para desenvolver, compartilhar e interoperar Sistemas de Organização de Conhecimento em ambientes web (CATARINO; CERVANTES; ANDRADE, 2015; SANTOS; MOREIRA, 2018). Nesse sentido, a representação do tesouro da UNESCO (PASTOR-SÁNCHEZ, 2016), o desenvolvimento de um tesouro no domínio da Ciência do Mar (MISOGUTI; RAMALHO, 2019) e a estruturação do Tesouro Brasileiro de Ciência da Informação (SANTOS; CERVANTES; FUJITA, 2018) são exemplos da utilização do SKOS para compartilhar elementos de conhecimento dos referidos domínios.

Cabe ressaltar que este trabalho é inspirado nos exemplos citados, principalmente, na contribuição do Tesouro Brasileiro de Ciência da Informação (PINHEIRO; FERREZ, 2014). Em face disso, como uma pesquisa aplicada, objetiva-se a implementação de um tesouro para o domínio da Ciência da Computação, mediante o uso do SKOS. Pressupõe-se que o desenvolvimento do pretendo tesouro contribui sobremaneira no (re)conhecimento

dos conceitos, na comunicação para com os profissionais da referida área de conhecimento (professores, pesquisadores e discentes) e na condução de pesquisas interdisciplinares.

Para tanto, ao apresentar os esforços despendidos na concepção do Tesauro da Ciência da Computação, além dessa seção introdutória, este artigo aborda: i) uma fundamentação teórica sobre tesauros como Sistemas de Organização de Conhecimento e o SKOS como modelo para estruturação de tesauros em ambientes web; ii) o percurso metodológico adotado para desenvolver o tesauro proposto; iii) a apresentação do Tesauro da Ciência da Computação e o modo de recuperação de informação implementado na Web de Dados; e iv) as considerações finais e sugestão de trabalhos futuros.

2 TESAUROS E SIMPLE KNOWLEDGE ORGANIZATION SYSTEM (SKOS) - UMA REVISÃO INTERDISCIPLINAR

Historicamente, no contexto dos Sistemas de Organização de Conhecimento, os tesauros são usados para sistematizar conteúdos de conhecimento, expressando e relacionando o vocabulário inerente a um domínio particular (SANTOS; MOREIRA, 2018). De forma complementar, aliado às novas tecnologias, os tesauros tornam o conteúdo de conhecimento formalizado, identificável e interoperável por máquinas, considerando a dinamicidade da infraestrutura da web (LARA, 2013). Considerando essas assertivas, atualmente, o desenvolvimento de um tesauro se configura como uma atividade interdisciplinar, principalmente, aliando disciplinas da Ciência da Computação e da Ciência da Informação. Ou seja, perante as bases constitutivas, a Ciência da Computação oferece a infraestrutura tecnológica para implementação e uso de tesauros (linguagens, modelos, repositórios, motores de busca, entre outros), ensejando as melhores práticas da Web Semântica (W3C, 2020a). Por outro lado, a Ciência da Informação se preocupa com o estabelecimento conceitual de formalismos, padrões e boas práticas, repercutindo nos procedimentos metodológicos e artefatos para organizar conteúdos de conhecimento na forma de tesauros. A inter-relação disciplinar discutida é representada na Figura 1.

Figura 1 –Tesauro e sua interdisciplinaridade



Pontualmente, como um sistema especializado de organização de conhecimento (PINHEIRO; FERREZ, 2014), um tesouro pode ser percebido como uma “lista de termos em linguagem natural, normalizados, preferenciais e organizados de modo conceitual, de acordo com regras terminológicas próprias e ligados entre si por relações hierárquicas ou semânticas” (CUNHA; CAVALCANTI 2008, p. 362). Em outras palavras, Stock e Stock (2015, p. 691) pontuam que tesouros são sistemas de organização de conhecimento dinâmicos, com a capacidade de adaptar sua estrutura e seus elementos frente às mudanças no domínio de conhecimento a ser representado. Neste sentido, Campos e Gomes (2006) explicitam que, pela dinamicidade em organizar os conceitos e as relações existentes entre os conceitos de um domínio, os tesouros contribuem nas tarefas de recuperação [de conteúdos de conhecimento especializados], de acordo com os objetivos de busca dos usuários.

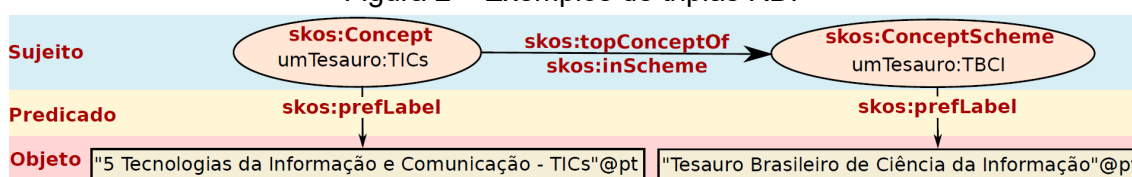
Esquemáticamente, os tesouros padronizam e descrevem os termos/conceitos e os organizam mediante três tipos de relações estruturantes: equivalência, hierárquica e associativa. Neste sentido, Austin e Dale (1993) formalizaram os elementos descritivos e estruturantes, como segue:

- **Termo Genérico (TG)** – relação hierárquica de sentido inverso para “Termo Específico”. Entre dois conceitos, indica aquele com conotação mais ampla. Por exemplo: o conceito “tesouro” tem o TG “Linguagens Documentárias” (PINHEIRO; FERREZ, 2014).
- **Termo Específico (TE)** - relação hierárquica de sentido inverso para “Termo Genérico”. Entre dois conceitos, indica aquele com conotação mais restritiva. Por exemplo: o conceito “tesouro” tem com um TE “microtesouro” (PINHEIRO; FERREZ, 2014).
- **Termo Relacionado (TR)** – relação associativa bidirecional entre dois conceitos. Por exemplo: “tesouro” é um TR a “sistemas de classificação” (PINHEIRO; FERREZ, 2014).
- **Usado Para (UP)** - relação associativa de sentido inverso para “USE”. Indica a existência de outro termo sinônimo, o qual é não é o preferido/adequado para rotular o conceito em análise. Por exemplo: o termo UP de “tecnologias da informação” é “tecnologia da informação e comunicação” (PINHEIRO; FERREZ, 2014).

- **USE** - relação associativa de sentido inverso para “Usado Para”. Indica a preferência de uso para rotular o conceito. Por exemplo: “tecnologias da informação e comunicação” é USE para “tecnologia da informação” (PINHEIRO; FERREZ, 2014).

Em face à vertente interdisciplinar apontada anteriormente (Ciência da Informação e Ciência da Computação) e considerando os recentes avanços tecnológicos para organizar e representar conhecimento, a estruturação de tesauros passou a incorporar importantes elementos da Web Semântica (LARA, 2013; CATARINO; CERVANTES; ANDRADE, 2015; RAMALHO, 2015; GABRIEL JUNIOR; LAIPELT, 2017; SANTOS; MOREIRA, 2018; SKOS, 2020). Especificamente, se refere ao SKOS, o qual provê um modelo descrito pelo *Framework* de Descrição de Recursos (*Resource Description Framework* - RDF) para estruturar o conteúdo de conhecimento em algum Sistema de Organização de Conhecimento (tesauros, taxonomias, folksonomias, vocabulários controlados, entre outros) na web (SKOS, 2020). A Figura 2 ilustra alguns recursos RDF (triplas sujeito, predicado e objeto) no contexto do uso do SKOS para representar elementos do Tesauro Brasileiro de Ciência da Informação.

Figura 2 – Exemplos de triplas RDF



Fonte: dados da pesquisa, baseando-se em Pinheiro e Ferrez (2014).

Estruturalmente, uma tripla RDF é composta por três partes: sujeito, predicado e objeto, sendo que um sujeito é relacionado a um objeto mediado um predicado. Considerando um exemplo abstraído da Figura 2, pode-se perceber uma tripla RDF em que um sujeito identificado por “umTesauro:TBCI” tem um predicado (*skos:prefLabel*) que aponta ao objeto cujo conteúdo é “Tesauro Brasileiro de Ciência da Informação”, expresso no idioma português (@pt). Outra tripla importante é a que vincula o sujeito “umTesauro:TICs” com o predicado *skos:inScheme* ao sujeito/objeto “umTesauro:TBCI”, denotando que o sujeito “umTesauro:TICs” está circunscrito a determinado *skos:ConceptScheme*. Ademais, o sujeito “umTesauro:TICs” figura como um recurso importante do sujeito/objeto “umTesauro:TBCI”, destacado pelo predicado *skos:topConceptOf*. Mais informações a respeito dos elementos que compõe o SKOS e outros exemplos são apresentados a seguir.

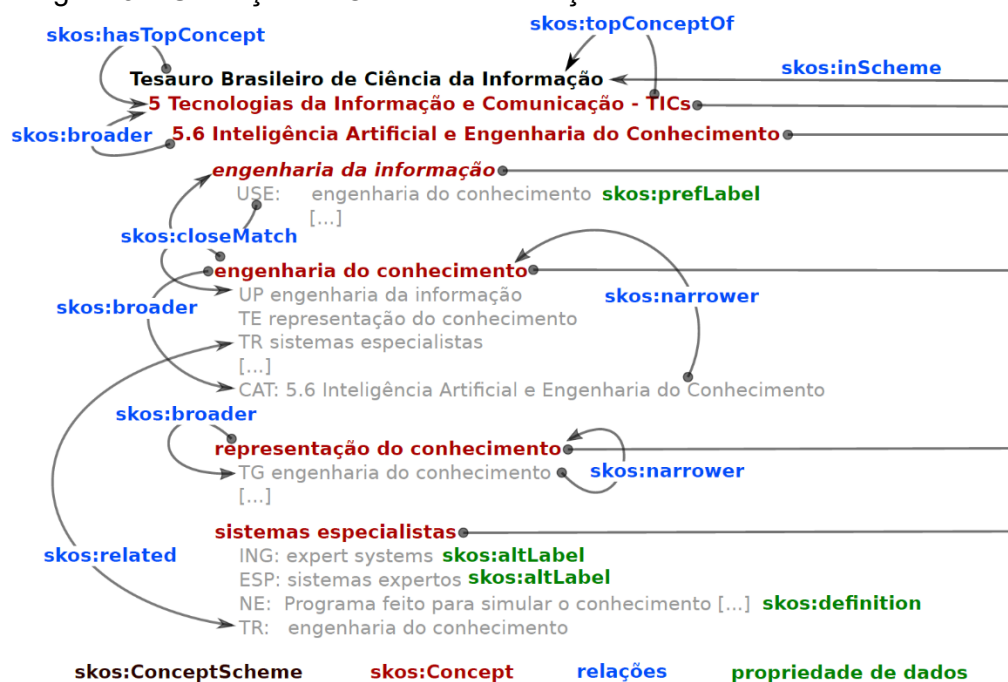
Cabe ressaltar que o SKOS é a recomendação do *World Wide Web Consortium* - W3C - (SKOS, 2020). Este destaca-se como o modelo adequado para interoperar os esquemas de organização e representação de conhecimento circunscritos nos ambientes da web (LARA, 2013). Neste trabalho, para manter a estrutura semântica de um tesauro, abstrai-se alguns recursos definidos no SKOS, sendo:

- ***skos:ConceptScheme*** – classe que agrega um conjunto de *skos:Concept* de um domínio particular.
- ***skos:Concept*** – classe que representa as unidades de conhecimento (uma ideia, conceito ou um termo) de um domínio.
- ***skos:hasTopConcept*** – relação que mapeia os *skos:Concept* mais genéricos (ou importantes), considerando um *skos:ConceptScheme*.
- ***skos:topConceptOf*** – relação inversa de *skos:hasTopConcept* que mapeia um *skos:Concept* mais genérico em um *skos:ConceptScheme*.
- ***skos:inScheme*** – relação que identifica quais *skos:Concept* pertencem a um *skos:ConceptScheme*.
- ***skos:broader*** – relação hierárquica entre dois *skos:Concept* ($A \rightarrow B$), representando que A tem em B um entendimento semanticamente mais amplo. Considerando a formalização de Austin e Dale (1993), *skos:broader* denota a relação “Termo Genérico - TG”.
- ***skos:narrower*** – relação hierárquica entre dois *skos:Concept* ($A \rightarrow B$). É a relação inversa de *skos:broader* e mapeia a especificidade de um conceito em relação a outro. Ou seja, representa que A tem em B uma especialização com entendimento mais restrito. Perante a formalização de Austin e Dale (1993), *skos:narrower* denota a relação “Termo Específico - TE”.
- ***skos:related*** – relação associativa entre dois *skos:Concept* ($A \leftrightarrow B$).
- ***skos:prefLabel*** – propriedade de dados que define o rótulo de um *skos:Concept* em linguagem natural.
- ***skos:altLabel*** – propriedade de dados que permite atribuir rótulos alternativos a um *skos:Concept*. Ressalta-se que o uso *skos:altLabel* pode ser uma alternativa para representar os rótulos de um *skos:Concept* em outros idiomas. Por exemplo, para um conceito cujo rótulo é “Ciência da Informação”@pt, a propriedade *skos:altLabel* poderia assumir os valores “*Information Science*”@en ou “*Ciencias de la Información*”@es para os idiomas inglês e espanhol, respectivamente.

- **skos:definition** – propriedade de dados que descreve um *skos:Concept*. Cabe ressaltar que, assim como *skos:altLabel*, para um conceito representado com o SKOS, pode-se apresentar várias definições, de acordo os idiomas de preferência (@pt, @en, @es, dentre outras).

Para melhor entender a relação do SKOS perante a organização de um tesauro, a Figura 3 ilustra os recursos SKOS (classes, relações e propriedades de dados) utilizados para representar alguns termos abstraídos do Tesauro Brasileiro de Ciência da Informação (PINHEIRO; FERREZ, 2014)

Figura 3 – Utilização SKOS frente às relações estruturantes de um tesauro



Fonte: dados da pesquisa, baseando-se em Pinheiro e Ferrez (2014) e SKOS (2020).

Considerando a figura citada, resumidamente, o Tesauro Brasileiro de Ciência da Informação torna-se um *skos:ConceptScheme* (sujeito), ao qual são vinculados (com o predicado *skos:inScheme*) os objetos (*skos:Concept*) rotulados por “5 - Tecnologia da Informação e Comunicação – TICs”, “5.6 – Inteligência Artificial e Engenharia do Conhecimento”, “engenharia da informação”, entre outros. No tesauro, hierarquicamente, “5.6 – Inteligência Artificial e Engenharia do Conhecimento” é um *skos:Concept*, melhor definido hierarquicamente pelos predicados *skos:hasTopConcept* e *skos:topConceptOf* entre os referidos recursos. Ao observar o sujeito denominado “engenharia do conhecimento”, pode-se perceber os predicados *skos:broader* e *skos:narrower* do referido sujeito para com o sujeito/objeto “5.6 – Inteligência Artificial e Engenharia do Conhecimento”, configurando outras formas para definir as relações de hierárquicas.

Quanto às propriedades de dados, ressalta-se que todos sujeitos são rotulados pelo predicado *skos:prefLabel*, subliminarmente. Em relação às demais propriedades de dados, ao observar o sujeito “sistemas especialistas”, é possível compreender que *skos:altLabel* é utilizado para rotular o sujeito em outros idiomas (espanhol e inglês, por exemplo). Ademais, ao sujeito “sistemas especialistas” uma definição é vinculada pelo predicado *skos:definition*.

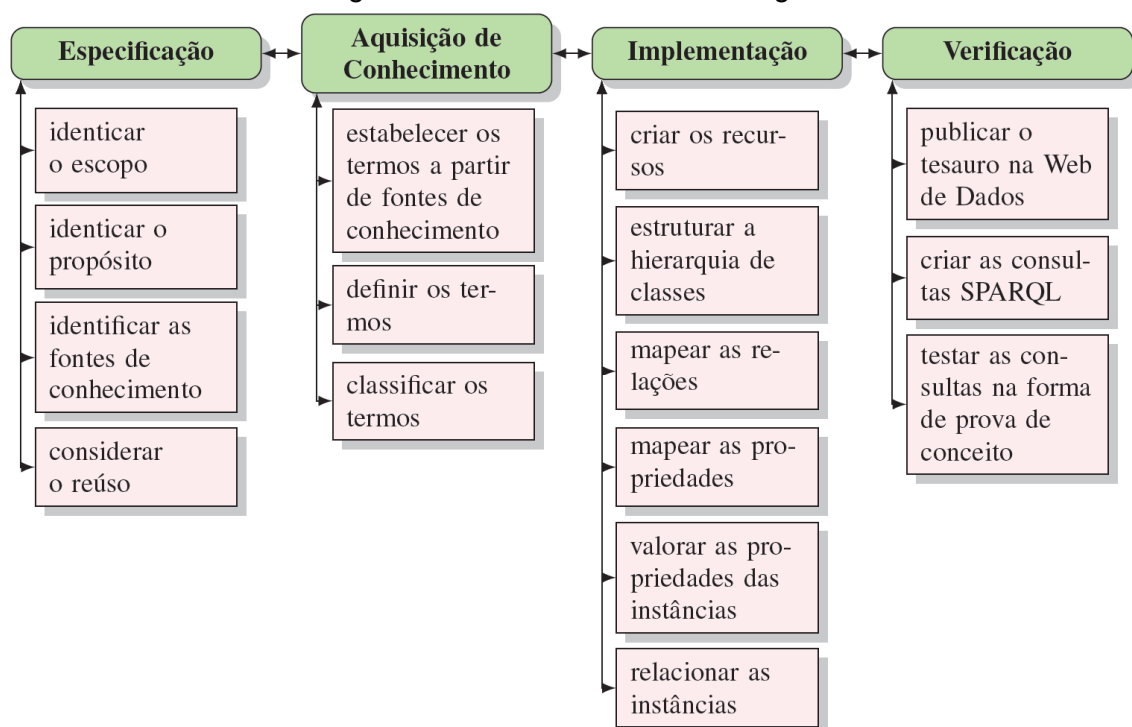
Mediante a Figura 3, considera-se que o exemplo ilustrado representa a interdisciplinaridade intrínseca do desenvolvimento de um Sistema de Organização de Conhecimento representado via os princípios da Web Semântica. Ao basear-se no exemplo, a próxima seção discorre sobre um percurso metodológico que permita desenvolver um trabalho similar, considerando a implementação de um Tesauro da Ciência da Computação.

3 PERCURSO METODOLÓGICO

O percurso metodológico adotado neste trabalho revela a natureza aplicada da pesquisa, principalmente, fundamentado em boas práticas para o desenvolvimento de sistemas de organização de conhecimento, conforme argumentos apresentados a seguir.

De modo geral, considera-se que a organização e a representação de conhecimento mediante a estruturação de tesauros é uma atividade próxima ao desenvolvimento de ontologias, por oportunizar o inter-relacionamento semântico de conceitos de um domínio específico (SALES; CAFÉ, 2008; SILVA; SOUZA; ALMEIDA, 2008; MACULAN; AGANETTE, 2017a; MACULAN; AGANETTE, 2017b; MOREIRA, 2019). Neste sentido, inspira-se no processo iterativo de desenvolvimento de ontologias adotado por Rautenberg (2009). Adequando o referido processo ao desenvolvimento de um tesauro (Figura 4), são estabelecidas as atividades de:

Figura 4 – Procedimento metodológico



Fonte: dados da pesquisa (2021).

1. **Especificação.** Tende-se a discernir sobre os custos do desenvolvimento do tesauro ao: (a) identificar o propósito - verificar em que ambiente o tesauro se insere, definir as circunstâncias do porquê deve ser desenvolvido, entre outros; (b) identificar o escopo – observar os requisitos de “quem são os usuários”, “quais são as intenções de uso”, entre outros; (c) identificar as fontes de conhecimento - procurar por livros, artigos, dicionários, entre outras fontes, dos quais pode-se abstrair conceitualizações; e (d) considerar o reúso - verificar a existência de ontologias ou vocabulários já estabelecidos para reutilizar conceitos.
2. **Aquisição do conhecimento.** Abstrai-se a maioria dos elementos de conhecimento aderentes ao contexto do tesauro. Iterativamente, consideram-se as tarefas de: (a) estabelecer os termos a partir das fontes de conhecimento - listar termos comumente utilizados no domínio, agregando os elementos interdisciplinares previamente estabelecidos em outros domínios; (b) classificar os termos - verificar na lista de termos se o entendimento de determinado recurso remete: a um conceito geral do domínio (classe); a uma associação entre os conceitos do domínio (relação); a um dado necessário, relativo a algum conceito do domínio (propriedade); ou a um exemplo de um conceito do domínio

(instância); e (c) definir os termos – atribuir uma definição conceitual aos termos selecionados (no contexto de um tesauro, é associar uma NE – nota explicativa).

3. **Implementação.** Utiliza-se os editores de ontologia, como por exemplo o Protégé (2020), para: (a) criar os recursos – a partir da lista de termos, os itens (classes, relações, propriedades e instâncias) são criados no arquivo inerente ao tesauro; (b) estruturar a hierarquia de classes – a partir da lista de classes, numa visão *top-down*, as classes com conceitos gerais (no contexto de um tesauro, é definir o TG – termo geral) são vinculadas às classes com conceitos semanticamente específicos (no contexto de um tesauro, é definir o TE – termo específico); (c) mapear as relações - para cada classe, agregam-se os termos tidos como “relação” e que evidenciam explicitamente o relacionamento da classe em questão para com as demais classes do domínio (no contexto de um tesauro, é definir o TR – termo relacionado); (d) mapear as propriedades - para cada classe, agregam-se os termos tidos como “propriedade de dados” e que são percebidos explicitamente como uma característica da classe em questão (no contexto de um tesauro, por exemplo, é a vinculação das propriedades ESP e ING); (e) valorar as propriedades das instâncias - definir os valores a cada propriedade de dados das instâncias criadas (por exemplo, o rótulo, os nomes alternativos e a definição – NE); e (f) relacionar as instâncias – inter-relacionar as instâncias do tesauro (no contexto de um tesauro, é valorar o TR entre as instâncias, definir o nome preferido – USE - ou os sinônimos - UP).

4. **Verificação.** Trata-se de uma atividade para averiguar se o tesauro atende aos requisitos estabelecidos na atividade “1. Especificação”, sendo que necessita-se: (a) publicar o tesauro na Web de Dados – para privilegiar a execução da prova de conceito pelos especialistas, o tesauro pode ser compartilhado em um *endpoint*¹ na Web de Dados; (b) criar as consultas SPARQL² - desenvolver consultas em linguagem SPARQL para recuperar os recursos codificados no tesauro em consonância aos requisitos previamente estabelecidos; e (c) testar as consultas na forma de prova de conceito – averiguar se as consultas SPARQL retornam os elementos requisitados.

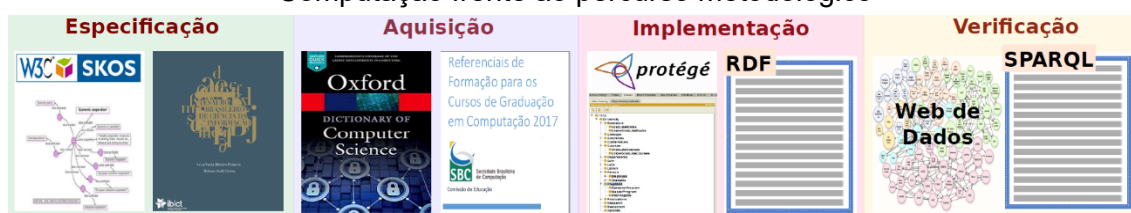
¹ Um endereço de Internet de acesso a um serviço da web de onde, mediante o uso de um protocolo específico de consulta, pode-se recuperar os recursos digitais em um formato específico (W3C, 2020c)

² Acrônimo recursivo para SPARQL *Protocol and RDF Query Language*, uma linguagem de consulta para dados representados no formato RDF (W3C, 2020b).

4 DESENVOLVIMENTO DO TESAURO PARA CIÊNCIA DA COMPUTAÇÃO

Em consonância ao percurso metodológico, o desenvolvimento do Tesauro da Ciência da Computação envolveu quatro atividades principais: Especificação, Aquisição, Implementação e Verificação.

Figura 5 – Principais recursos utilizados no desenvolvimento do Tesauro da Ciência da Computação frente ao percurso metodológico



Fonte: dados da pesquisa (2021).

A Figura 5 ilustra o referido percurso, evidenciando os principais recursos utilizados. Os detalhes do desenvolvimento do tesauro são abordados a seguir.

4.1 Especificação

As principais tarefas desempenhadas durante a atividade de “1. Especificação” circunscreveram a identificação das fontes de conhecimento aderentes ao Tesauro da Ciência da Computação. Alternativamente, também pode-se considerar o reúso de outros Sistemas de Organização de Conhecimento. Neste sentido, cinco fontes de conhecimento foram identificadas:

- **Butterfield e Ngondi (2016)** – intitulado *A Dictionary of Computer Science*, é uma obra importante da *Oxford University Press* que agrega cerca de 6.500 verbetes das disciplinas da Ciência da Computação.
- **Pinheiro e Ferrez (2014)** – refere-se ao Tesauro Brasileiro de Ciência da Informação, sendo um exemplo inspirador para organizar tesauros de outros domínios.
- **SKOS (2020)** - denominado *Simple Knowledge Organization System Reference*, é o sítio do *World Wide Web Consortium* em que se pode verificar os elementos constituintes e atualizações do SKOS.

- **Wikipedia (2020)** – sítio na web do qual é possível abstrair as conceitualizações temporárias não registradas em Butterfield e Ngondi (2016) e que merecem serem reeditadas no futuro; e
- **Zorzo et al. (2017)** – documento que expressa os Referenciais de Formação para os Cursos de Graduação em Computação publicado pela Sociedade Brasileira de Computação - SBC. São definidos os referenciais para os cursos de Ciência da Computação, Engenharia da Computação, Engenharia de Software, Licenciatura em Computação, Sistemas de Informação e Cursos Superiores em Computação. A partir deste documento pôde-se perceber os termos que constituem o tesouro proposto e a forma de relacionar os conceitos existentes.

4.2 Aquisição do conhecimento

Na atividade “2. Aquisição do conhecimento”, a partir das fontes de conhecimento identificadas, prioritariamente, os termos aderentes ao domínio em investigação devem ser enumerados e classificados, abstraindo também as definições e relações entre os conceitos. Neste sentido, ao considerar o SKOS como um modelo para Sistemas de Organização de Conhecimento, o SKOS contribui à aquisição do conhecimento ao permitir a especialização de elementos de conhecimento presentes em um domínio (GABRIEL JUNIOR; LAIPELT, 2017). Segundo os autores, o SKOS possibilita a integração de novos elementos (classes, relações e propriedades de dados) para melhor circunscrever o conhecimento do domínio. Dito isso e a partir dos Referenciais de Formação para os Cursos de Graduação em Computação (ZORZO et al., 2017), alguns elementos foram abstraídos para melhor caracterizar os termos comumente utilizados no contexto do Tesouro da Ciência da Computação, sendo eles:

- **sbc:Eixo** – especialização do *skos:Concept* que representa os diferentes eixos de formação de um cientista da computação. Cada eixo deve relacionar os conteúdos de conhecimento basilares ao desenvolvimento das competências profissionais. Como exemplo, destaca-se o eixo de formação denominado “Eixo 2 – Desenvolvimento de Sistemas”.
- **sbc:CompetenciaEixo** – especialização do *skos:Concept* que representa o conjunto de competências associadas a algum eixo de formação. Por exemplo, como competências para o Eixo 2 destacado anteriormente, os egressos devem ser capazes de “desenvolver sistemas computacionais que atendam a qualidade

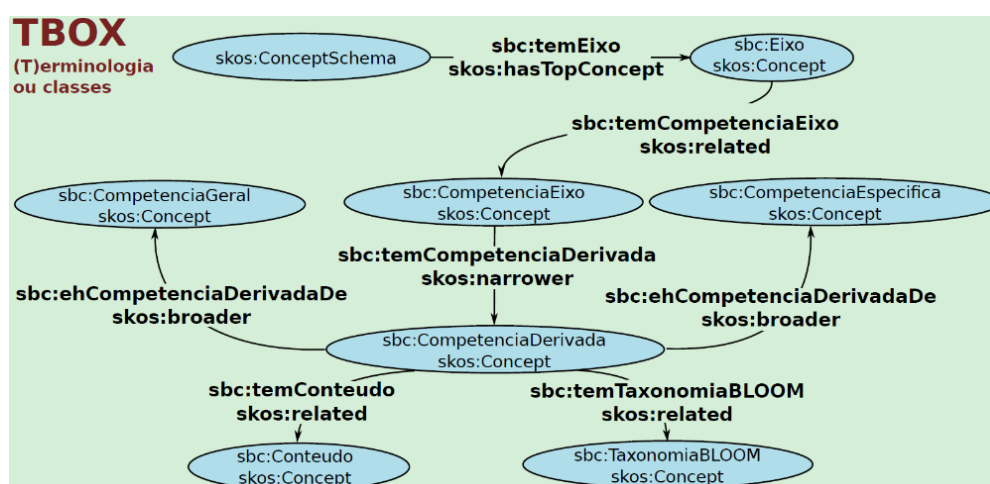
de processo e de produto, considerando princípios e boas práticas de engenharia de sistemas e engenharia de software [...]”.

- **sbc:CompetenciaGeral** – especialização do *skos:Concept* que representa as competências e/ou habilidades gerais de um egresso, como por exemplo “CG-III. Resolver problemas usando ambientes de programação”.
- **sbc:CompetenciaDerivada** – especialização do *skos:Concept* que mapeia a materialização conjunta de uma *sbc:CompetenciaEixo* e uma *sbc:CompetenciaGeral* em função dos *sbc:Conteudo* a serem ministrados. Por exemplo, na *sbc:CompetenciaDerivada* denominada “C.2.1. Resolver problemas usando ambientes de programação (CG-III)”, a parte do rótulo “C.2.1” identifica a primeira competência derivada das “Competências do Eixo II” e a parte do rótulo “(CG-III)” vincula a competência derivada à *sbc:CompetenciaGeral* “Competência Geral III”. Em suma, o exemplo denota que, para um egresso desenvolver sistemas, ele deve ser capacitado ao uso de ambientes de programação.
- **sbc:CompetenciaEspecific**a – especialização do *skos:Concept* que representa as competências específicas desenvolvidas em um Bacharelado em Ciência da Computação, as quais um egresso pode ser capacitado a atender. Como exemplo, destaca-se a *sbc:CompetenciaEspecific*a “Identificar e gerenciar os riscos que podem estar envolvidos na operação de equipamentos de computação (incluindo os aspectos de dependabilidade e segurança) (CE-III)”.
- **sbc:Conteudo** – especialização do *skos:Concept* que representa os conhecimentos que devem ser ministrados em um Bacharelado em Ciência da Computação, visando o desenvolvimento das competências.
- **sbc:TaxonomiaBLOOM** – especialização do *skos:Concept* que tipifica uma *sbc:CompetenciaDerivada* de acordo com seis níveis incrementais de aquisição de conhecimento: lembrar, entender, aplicar, analisar, avaliar e criar.
- **sbc:temEixo** – especialização da relação *skos:hasTopConcept* que relaciona o *skos:ConceptScheme* “Tesouro da Ciência da Computação” a seus conceitos mais genéricos considerados como *sbc:Eixo*.
- **sbc:temCompetenciaEixo** – especialização da relação *skos:related* que associa um *sbc:Eixo* a suas *sbc:CompetenciaEixo*.
- **sbc:temCompetenciaDerivada** – especialização da relação *skos:narrowed* que especifica as *sbc:CompetenciaDerivada* de uma *sbc:CompetenciaEixo*.

- **sbc:ehCompetenciaDerivadaDe** – especialização da relação *skos:broader* que especializa uma *sbc:CompetenciaDerivada* a partir de uma *sbc:CompetenciaGeral* ou de uma *sbc:CompetenciaEspecificas*.
- **sbc:temConteudo** – especialização da relação *skos:related* que determina quais *sbc:Conteudo* são associados a uma *sbc:CompetenciaDerivada*; e
- **sbc:temTaxonomiaBLOOM** - especialização da relação *skos:related* que atribui a respectiva *sbc:TaxonomiaBLOOM* para uma *sbc:CompetenciaDerivada*.

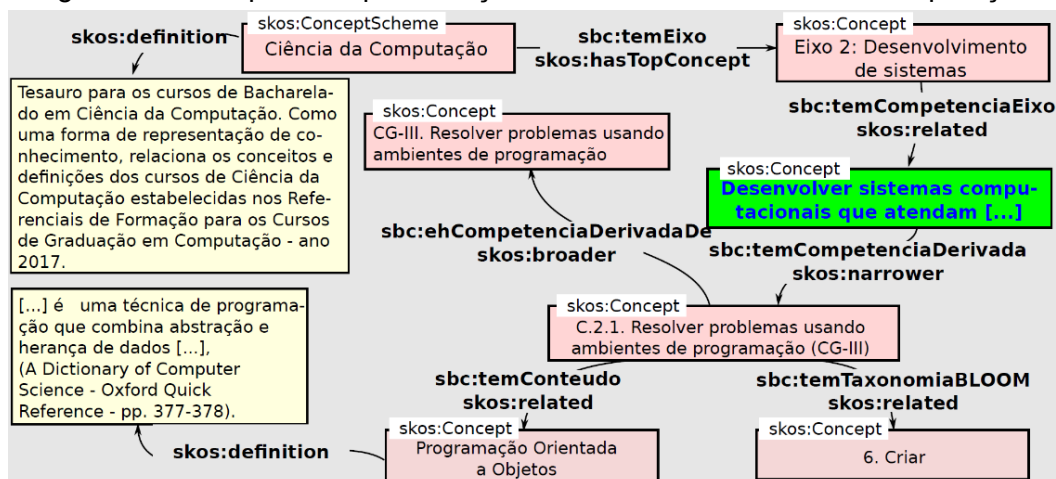
Considerando os termos relacionados anteriormente, auxiliando no entendimento do Sistema de Organização de Conhecimento em formação, a Figura 6 ilustra como se configura a representação semântica do Tesouro da Ciência da Computação aliada ao SKOS.

Figura 6 – Especialização do SKOS frente os termos dos Referenciais de Formação para os Cursos de Graduação em Computação – Bacharelado em Ciência da Computação



Fonte: dados da pesquisa (2021).

Figura 7 – Exemplo de representação do tesauro da Ciência da Computação



Fonte: dados da pesquisa (2021).

Já a Figura 7 exemplifica uma representação concreta utilizando o Tesauro da Ciência da Computação considerando o inter-relacionamento dos elementos conceituais vinculados à competência de eixo “desenvolvimento de sistemas computacionais [...]” criando soluções mediante ambientes computacionais e os conteúdos da Programação Orientada a Objetos.

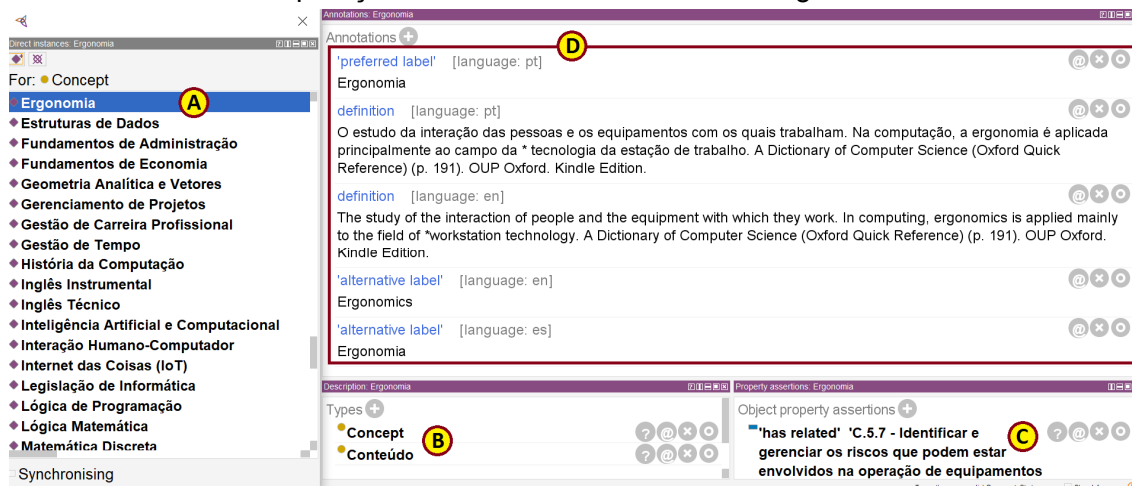
4.3 Implementação

Como mencionado, na atividade “3. Implementação” utiliza-se os editores de ontologia como o Protégé (2020) para editar o arquivo de um tesauro. Na atividade de edição, principalmente: (a) os recursos (classes, relações, propriedades e instâncias) são criados; (b) a estrutura hierárquica do tesauro é definida; (c) as associações entre os recursos são mapeadas; e (d) as instâncias são valoradas.

Para ilustrar o uso do ambiente Protégé, a Figura 8 exemplifica como ocorre a instanciação do conceito “Ergonomia” no Tesauro da Ciência da Computação (Figura 8.A).

Pontualmente, ao observar a Figura 8.B, o conceito é classificado, ao mesmo tempo, como *skos:Concept* e *sbc:Conteudo*, sendo associado à competência derivada (Figura 8.C) “C.5.7 - Identificar e gerenciar os riscos que podem estar envolvidos [...]”. Ademais, a Figura 8.D ilustra as propriedades de dados inerentes ao conceito, sendo destacados os idiomas utilizados para nomeação e definição (neste caso, @en, @es e @pt).

Figura 8 – Representação do ambiente utilizado no desenvolvimento do Tesauro da Ciência da Computação – evidenciando o conceito “Ergonomia”

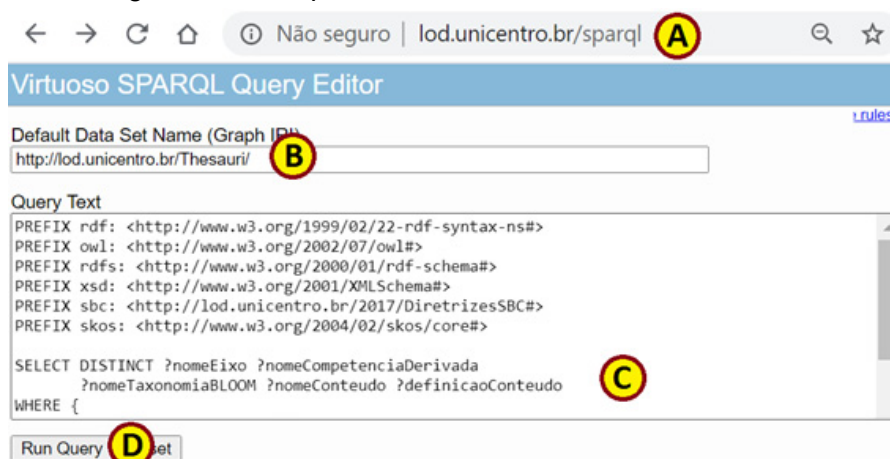


Fonte: dados da pesquisa (2021).

4.4 Verificação

Com o objetivo de verificar a consistência do tesauro desenvolvido, uma versão do Tesauro da Ciência da Computação foi publicada na Web de Dados. Neste sentido, a Figura 9 apresenta o sítio em que consultas ao referido Sistema de Organização do Conhecimento podem ser realizadas (<http://lod.unicentro.br/sparql>). Na Figura são destacados: (A) o endereço do *endpoint* que mantém o tesauro; (B) o URI de todos os tesouros mantidos no *endpoint*; (C) a área de texto para desenvolvimento de consultas SPARQL; e (D) a operação para execução de consultas SPARQL.

Figura 9 – Sítio para consulta do tesauro desenvolvido



Fonte: dados da pesquisa (2021).

Para privilegiar a prova de conceitos, uma consulta geral ao Tesauro da Ciência da Computação foi desenvolvida em linguagem SPARQL (Listagem 1), permitindo averiguar se o tesauro apresenta os elementos requisitados.

Listagem 1 – Exemplo de consulta SPARQL ao Tesauro desenvolvido

```

01 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
02 PREFIX owl: <http://www.w3.org/2002/07/owl#>
03 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
04 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
05 PREFIX sbc: <http://lod.unicentro.br/2017/DiretrizesSBC#>
06 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
07
08 SELECT DISTINCT ?nomeEixo ?nomeCompetenciaDerivada
09 ?nomeTaxonomiaBLOOM ?nomeConteudo ?definicaoConteudo
10 WHERE {
11   ?tesouroCC rdf:type skos:ConceptScheme .
12   ?eixo rdf:type skos:Concept .
13   ?competenciaEixo rdf:type skos:Concept .
14   ?competenciaDerivada rdf:type skos:Concept .
15   ?taxonomiaBLOOM rdf:type skos:Concept .
16   ?conteudo rdf:type skos:Concept .
17
18   ?eixo rdf:type sbc:Eixo .
19   ?competenciaEixo rdf:type sbc:CompetenciaEixo .
20   ?competenciaDerivada rdf:type sbc:CompetenciaDerivada .
21   ?taxonomiaBLOOM rdf:type sbc:TaxonomiaBLOOM .
22   ?conteudo rdf:type sbc:Conteudo .
23
24   ?eixo skos:inScheme sbc:CC .
25   ?competenciaEixo skos:inScheme sbc:CC .
26   ?competenciaDerivada skos:inScheme sbc:CC .
27   ?taxonomiaBLOOM skos:inScheme sbc:CC .
28   ?conteudo skos:inScheme sbc:CC .
29
30   ?eixo skos:related ?competenciaEixo .
31   ?competenciaEixo skos:narrower ?competenciaDerivada .
32   ?competenciaDerivada skos:related ?conteudo .
33   ?competenciaDerivada skos:related ?taxonomiaBLOOM .
34   ?eixo skos:prefLabel ?nomeEixo .
35   ?competenciaDerivada skos:prefLabel ?nomeCompetenciaDerivada .
36   ?taxonomiaBLOOM skos:prefLabel ?nomeTaxonomiaBLOOM .
37   ?conteudo skos:prefLabel ?nomeConteudo .
38   OPTIONAL {
39     ?conteudo skos:definition ?definicaoConteudo .
40     FILTER (LANGMATCHES(LANG(?definicaoConteudo), "pt"))
41     #FILTER (LANGMATCHES(LANG(?definicaoConteudo), "en"))
42   }
43 } ORDER BY ?competenciaDerivada ?conteudo

```

Fonte: dados da pesquisa (2021).

Cabe ressaltar que a consulta SPARQL listada pode ser customizada em consonância à outras requisições de usuários. Ademais, a consulta também pode ser executada a partir do URL encurtado <https://tinyurl.com/ycqszn7u>.

5 CONSIDERAÇÕES FINAIS

Este artigo disserta sobre o esforço inicial dispendido para desenvolver um tesauro. Utilizando a base teórica Sistema de Organização de Conhecimento, desenvolveu-se uma versão do Tesauro para a Ciência da Computação. No desenvolvimento, utilizou-se o ambiente Protégé, estendendo o Sistema de Organização de Conhecimento conhecido

como SKOS – *Simple Knowledge Organization System*. Destaca-se que o tesauro é inspirado principalmente na iniciativa do Tesauro Brasileiro de Ciência da Informação e nos Referenciais de Formação para os Cursos de Graduação em Computação, sendo disponibilizado na Web de Dados a partir do URL encurtado <https://tinyurl.com/ycqszn7u>.

Admite-se que o público-alvo do Tesauro da Ciência da Computação são profissionais, professores, pesquisadores, alunos de graduação/pós-graduação vinculados aos conteúdos e pesquisas interdisciplinares, principalmente, para subsidiar a comunicação científica entre os referidos atores. Por tratar-se de uma pesquisa aplicada, observou-se que os processos para o desenvolvimento de ontologias também podem ser empregados para o desenvolvimento de tesauros. Ademais, a utilização do SKOS como modelo para o desenvolvimento de tesauros mostrou-se adequado, conforme as premissas deste sistema na organização e na representação de recursos digitais.

Com base no estudo realizado, futuramente, almeja-se enriquecer o tesauro desenvolvido com novos recursos informacionais do domínio da Ciência da Computação, os quais não foram contemplados na versão atual. Entende-se que a atualização do tesauro é uma atividade perene no tempo, visto que o conjunto de termos circunscritos à área é renovado constantemente. Também pretende-se atuar na melhor especificação dos conteúdos da área, explicitando os tópicos intrínsecos a serem abordados.

REFERÊNCIAS

- AUSTIN, D.; DALE, P. **Diretrizes para o estabelecimento e desenvolvimento de tesauros monolíngües**. Brasília: Ibict/SENAI, 1993. *E-book*. Disponível em: https://livroaberto.ibict.br/bitstream/1/731/7/Diretrizes_estabelecimento_tesauros.pdf. Acesso em: 21 ago. 2021.
- BRÄSCHER, M.; CAFÉ, L. Organização da Informação ou Organização do Conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 8, 2008, São Paulo. **Anais [...]**. São Paulo: USP, 2008.
- BUTTERFIELD, A.; NGONDI, G. E. **A Dictionary of Computer Science**. 7ª ed. New York: Oxford University Press, 2016.
- CAMPOS, M. L. A., GOMES, H. E. Metodologia de elaboração de tesauro conceitual: a categorização como princípio norteador. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.11, n. 3, p. 348-359, 2006.
- CATARINO, M. E.; CERVANTES, B. M. N.; ANDRADE, I. A. de. A Representação Temática no Contexto da Web Semântica. **Informação & Sociedade: Estudos**, João Pessoa, v. 25, n. 3, p. 105-116, 2015.
- CUNHA, M. B. da; CAVALCANTI, C. R. de O. **Dicionário de biblioteconomia e arquivologia**. Brasília: Briquet de Lemos, 2008.
- FERREIRA, A. C.; MACULAN, B. C. M; dos S. Metodologias para Revisão e Atualização de Tesauros: mapeamento da literatura. **Informação & Informação**, Londrina, v. 25, n. 1, p. 229–253, 2020.

- GABRIEL JUNIOR, R. F.; LAIPELT, R. do C, F. THESA: ferramenta para construção de tesouro semântico aplicado interoperável. **P2P & INOVAÇÃO**, Rio de Janeiro, v. 3, n. 2, p.124-145, 2017.
- LARA, M. L. G. de. Linguagens documentárias e sistemas de organização do conhecimento no contexto da web semântica. **TransInformação**, Campinas, v. 25, n. 2, p. 145-150, 2013.
- MACULAN, B. C. M. dos S.; AGANETTE, E. C. Definições: Convergência na Construção de Tesouros e Ontologias. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 18., 2017, Marília. **Anais [...]**. Marília: UNESP, 2017a.
- MACULAN, B. C. M. dos S.; AGANETTE, E. Desambiguação de relações em tesouros e o seu reuso em ontologias. **Ciência da Informação**, Brasília, v.46, n.1, p.102-119, 2017b.
- MISOGUTI, D.; RAMALHO, R. A. S. Modelo SKOS para controle do vocabulário de Ciências do Mar. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 20., 2019, Florianópolis. **Anais [...]**. Florianópolis: UFSC, 2019.
- MOREIRA, W. Tesouros e ontologias como modelos de sistemas de organização do conhecimento. **Brazilian Journal of Information Studies: Research Trends**, Marília, v. 13, n. 1, p.15-20, 2019.
- PASTOR-SÁNCHEZ, J.-A. Proposal to represent the UNESCO Thesaurus for the Semantic Web applying ISO-25964. **Brazilian Journal of Information Studies: Research Trends**, Marília, v. 10, n. 1, p. 1-8, 2016.
- PINHEIRO, L. V. R.; FERREZ H, D. **Tesouro Brasileiro de Ciência da Informação**. Rio de Janeiro; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), 2014. E-book. Disponível em: <https://ibict.br/images/internas/TESAURO-COMPLETO-FINAL-COM-CAPA- 24102014.pdf>. Acesso em: 17 mai. 2021.
- PROTÉGÉ**. Disponível em: <https://protege.stanford.edu/>. Acesso em: 14 ago. 2020.
- RAMALHO, R. A. S. Análise do Modelo de Dados SKOS: Sistema de Organização do Conhecimento Simples para a Web. **Informação & Tecnologia**, João Pessoa, v. 2, n. 1, p. 66-79, 2015.
- RAUTENBERG, S. **Modelo de conhecimento para mapeamento de Instrumentos da gestão do conhecimento e de agentes computacionais da engenharia do conhecimento baseado em ontologias**. 2009. Tese (Doutorado em Engenharia e Gestão do Conhecimento) – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2009.
- SALES, R. de; CAFÉ, L. Semelhanças e Diferenças entre Tesouros e Ontologias. **DataGramaZero**, João Pessoa, v. 9 n. 4, 2008.
- SANTOS, J. C. F. dos; CERVANTES, B. M. N.; FUJITA, M. S. L. Tesouro Eletrônico: importação no Tematres e disponibilização na web. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 19., 2018, Londrina. **Anais [...]**. Londrina: UEL, 2018.
- SANTOS, J. C. F. dos; MOREIRA, W. SKOS: uma análise sobre as abordagens e suas as aplicações na Ciência da Informação. **Informação & Informação**, Londrina, v. 23, n. 3, p. 362–389, 2018.
- SILVA, D. L. da; SOUZA, R. R.; ALMEIDA, M. B. Ontologias e vocabulários controlados: comparação de metodologias para construção. **Ciência da Informação**, Brasília, v. 37, n. 3, p. 60-75, 2008.
- SKOS. **SKOS Simple Knowledge Organization System Reference**. Disponível em: <https://www.w3.org/TR/2009/REC-skos-reference-20090818/>. Acesso em: 08 ago. 2020.
- STOCK, W. G.; STOCK, M. **Handbook of Information Science**. Berlin, Boston: De Gruyter Saur, 2015.
- W3C. **Data on the Web best practices: W3C recommendation. 31 January 2017**. Disponível em: <https://www.w3.org/TR/2017/REC-dwbp-20170131/>. Acesso em: 09 set. 2020a.

W3C. **SPARQL Query Language for RDF**. Disponível em: <https://www.w3.org/TR/rdf-sparql-query/>. Acesso em: 09 set. 2020b.

W3C. **Web Services Description Requirements**. Disponível em: <https://www.w3.org/TR/2002/WD-ws-desc-reqs-20021028/#normDefs>. Acesso em: 09 set. 2020c.

WIKIPEDIA. **Wikipédia, a enciclopédia livre**. Disponível em: https://pt.wikipedia.org/wiki/Wikip%C3%A9dia:P%C3%A1gina_principal. Acesso em: 07 dez. 2020.

ZORZO, A. F.; NUNES, D.; MATOS, E.; STEINMACHER, I.; LEITE, J.; ARAUJO, R. M.; CORREIA, R.; MARTINS, S. **Referenciais de Formação para os Cursos de Graduação em Computação**. Porto Alegre: Sociedade Brasileira de Computação, 2017. E-book. Disponível em: <https://www.sbc.org.br/documentos-da-sbc/send/127-educacao/1155-referenciais-de-formacao-para-cursos-de-graduacao-em-computacao-outubro-2017>. Acesso em: 17 mai. 2021.

NOTAS

AGRADECIMENTOS

O primeiro autor agradece à Fundação Araucária pelo suporte financeiro ao projeto de pesquisa intitulado “Curadoria Digital e Dados Abertos Conectados: um estudo da preservação de recursos digitais na Web de Dados para estudos cientométricos” (FA - Convênio 49773/19FA).

CONTRIBUIÇÃO DE AUTORIA

Concepção e elaboração do manuscrito: Rautenberg, S.; Souza, L. de.

Coleta de dados: Kelniar, J. P.

Análise de dados: Rautenberg, S.; Kelniar, J. P.

Discussão dos resultados: Rautenberg, S.; Souza, L. de; Kelniar, J. P.

Revisão e aprovação: Rautenberg, S.; Souza, L. de.

CONJUNTO DE DADOS DE PESQUISA

O conjunto de dados que dá suporte aos resultados deste estudo não está disponível publicamente.

FINANCIAMENTO

Fundação Araucária (FA - Convênio 49773/19FA).

CONSENTIMENTO DE USO DE IMAGEM

Não se aplica

APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA

Não se aplica

CONFLITO DE INTERESSES

Não se aplica

LICENÇA DE USO

Os autores cedem à **Encontros Bibli** os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution](#) (CC BY) 4.0 International. Esta licença permite que **terceiros** remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os **autores** têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

PUBLISHER

Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES

Enrique Muriel-Torrado, Edgar Bisset Alvarez, Camila Barros.

HISTÓRICO

Recebido em: 30-03-2021 – Aprovado em: 28-05-2021 – Publicado em: 11-06/2021

