



XXI ENANCIB

Encontro Nacional de Pesquisa em Ciência da Informação

50 anos de Ciência da Informação no Brasil:
diversidade, saberes e transformação social

Rio de Janeiro • 25 a 29 de outubro de 2021

XXI Encontro Nacional de Pesquisa em Ciência da Informação – XXI ENANCIB

GT-8 – Informação e Tecnologia

MAPEAMENTO DE TAXONOMIAS NO DOMÍNIO CIENTÍFICO: alinhando recursos para internacionalização da Plataforma IAraucária

MAPPING SCIENTIFIC TAXONOMIES: resource alignment for IAraucária Platform internationalization

Sandro Rautenberg - Universidade Estadual do Centro-Oeste (UNICENTRO)
srautenberg@unicentro.br

Weber Claudio Francisco Nunes da Silva - Universidade Estadual do Centro-Oeste (UNICENTRO)
wwwclaudion@gmail.com

Luiz Márcio Spinosa - Fundação Araucária
spinosa@fundacaoaraucaria.org.br

Modalidade: Trabalho Completo

Resumo: Encontrar informação relevante é uma atividade custosa e complexa que pode ser mediada por vocabulários controlados, como as taxonomias. O artigo objetiva discutir o uso de taxonomias para classificação de áreas de conhecimento em um sistema de localização de especialistas. Metodologicamente, utiliza-se o *Linked Data Lifecycle* para integrar duas taxonomias. Como resultado, a Tabela de Áreas de Conhecimento CNPq e a lista *All Scientific Journals Codes* são integradas com o *Simple Knowledge Organization System* - SKOS, constituindo: i) uma contribuição ao sistema de localização de especialistas denominado Plataforma IAraucária; e ii) um suporte a outros sistemas de informação, haja vista que as taxonomias são compartilhadas na Web de Dados.

Palavras-Chave: Representação de Conhecimento; Taxonomias; *Simple Knowledge Organization System* - SKOS; Plataforma IAraucária.

Abstract: Finding relevant information is a complex activity that can be mediated by controlled vocabularies such as taxonomies. This paper discusses the use of taxonomies for classifying the knowledge fields in an expertise locator system. Methodologically, *Linked Data Lifecycle* is used to integrate two taxonomies. As a result, the *Tabela de Áreas de Conhecimento CNPq* and the *All Scientific Journals Codes* list are integrated by using the *Simple Knowledge Organization System* - SKOS, establishing: **i)** a contribution to the expertise locator system named *Plataforma IAraucária*; and **ii)** a support to other information systems, given that the taxonomies are shared on the *Web of Data*.

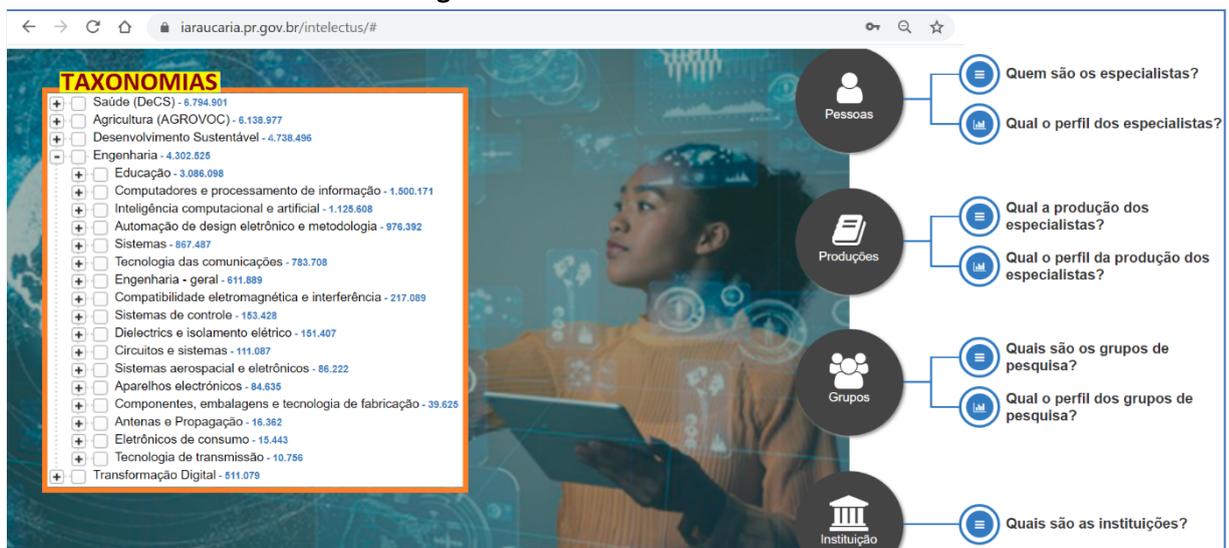
Keywords: Knowledge Representation; Taxonomies; *Simple Knowledge Organization System* - SKOS; *Plataforma IAraucária*.

1 INTRODUÇÃO

A organização e a representação de informação e conhecimento são temas interdisciplinares e relevantes à Ciência da Informação, sendo suas fronteiras expandidas pelos avanços tecnológicos, como por exemplo, os padrões, modelos e melhores práticas da Web Semântica (W3C, 2021a; W3C, 2021b). Neste sentido, os recursos informacionais são produzidos e disponibilizados em direção aos benefícios de reuso, compreensão, inter-relacionamento, descoberta, confiança, acesso, interoperabilidade e processabilidade desses recursos nos ambientes da web (LÓSCIO; BURLE; CALEGARI, 2021). Isso enseja desafios e oportunidades de pesquisa para a utilização de informação e conhecimento, mediante o uso de vocabulários controlados na organização e na representação da informação, auxiliando a exploração de recursos na Internet.

Considerando as assertivas anteriores, conceitualmente, este artigo delimita-se ao uso de uma classe especializada de vocabulários controlados, as taxonomias, para representar informação na Web Semântica. De forma geral, como objeto de pesquisa, as taxonomias são utilizadas para rotulagem, tradução, indicação de associações, navegação e recuperação de recursos em sistemas informacionais (NISO, 2010, p. 11). Em se tratando de uma pesquisa aplicada, o *locus* da pesquisa é pautado em uma das ações para o fomento à pesquisa paranaense, a consolidação de uma solução sociotécnica (Figura 1) denominada Plataforma Iaraucária (IARAUCÁRIA, 2021).

Figura 1: Plataforma Iaraucária.



Fonte: Adaptada (IARAUCÁRIA, 2021).

Em poucas palavras, a plataforma IAraucária oferece recursos computacionais para auxiliar a consolidação de NAPIs - Novos Arranjos de Pesquisa e Inovação - uma rede de pesquisadores e profissionais para cocriar e coproduzir conhecimento no âmbito da Ciência, Tecnologia e Inovação no território paranaense (FUNDAÇÃO ARAUCÁRIA, 2021). Com a plataforma, as competências de pesquisadores são classificadas mediante as taxonomias que guiam as buscas especializadas em uma base de dados de currículos Lattes. Dentre as taxonomias reconhecidas e utilizadas na plataforma, pode-se citar a AGROVOC (FAO, 2021) e a DeCS (PELLIZZON, 2004), por exemplo. Em resumo, para amparar a formação de um NAPI, a Plataforma IAraucária atua como um sistema de localização de especialistas, identificando os colaboradores em consonância às competências técnicas e científicas requeridas em uma demanda do território.

Com o intuito de contribuir no desenvolvimento da Plataforma IAraucária, pressupõe-se que o uso de outras taxonomias aprimora o sistema de localização de especialistas. Principalmente, ao que se refere o uso de vocabulários controlados para melhorar a classificação de especialistas ao relacionar o domínio de conhecimento de projetos e produções científicas registrados perante as demandas científico-tecnológicas em análise.

Diante disso, na forma de um estudo de caso, objetiva-se investigar a integração de taxonomias para classificação de domínios do conhecimento para, futuramente, expandir a forma de recuperação de recursos informacionais na Plataforma IAraucária.

Para versar a respeito dos esforços realizados, além dessa seção introdutória, o artigo compreende: **i)** a fundamentação teórica, a qual aborda o conceito de taxonomias e o modelo da Web Semântica denominado *Simple Knowledge Organization System* – SKOS; **ii)** o procedimento metodológico que amparou a evolução das taxonomias em direção aos preceitos da Web Semântica; **iii)** o resultado alcançado na forma do registro do percurso realizado durante o desenvolvimento; **iv)** a discussão da contribuição do trabalho realizado; e **vi)** as considerações finais, com o direcionamento de trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este artigo é circunscrito à realidade de um tipo especializado de vocabulário controlado, as taxonomias, instanciando-as segundo os preceitos de um modelo de organização e representação de informação da Web Semântica, o SKOS (W3C, 2021a). Em

face disso, a seguir são tecidos apontamentos sobre vocabulários controlados, taxonomias e SKOS.

Segundo a NISO (2010, p. 5), um vocabulário controlado constitui uma representação de um conjunto finito de termos em consonância ao entendimento conceitual e limite de um domínio de interesse. Ressalta-se que, para cada termo presente no domínio em investigação, deve-se ter um entendimento inequívoco e não redundante para: **i)** evitar a ambiguidade; **ii)** controlar a sinonímia existente; **iii)** estabelecer as relações entre os termos; e **iv)** realizar as validações do vocabulário perante o domínio de conhecimento. Em sua essência, os vocabulários controlados listam e organizam recursos informacionais de um domínio, ao associar termos significantes e semântica aos objetos representados. Dito isso, os vocabulários controlados servem a cinco propósitos (NISO, 2010, p. 11):

- **Tradução.** Fornece um meio para sistematizar a linguagem natural de um domínio de conhecimento em um vocabulário adequado para indexação e recuperação de recursos em sistemas de informação.
- **Consistência.** Promove a uniformidade no formato e na atribuição dos termos ao organizar, classificar e representar o conhecimento de um domínio.
- **Indicação de associações.** Indica as relações semânticas (hierárquicas ou transitivas) entre os termos.
- **Rotulagem e Navegação.** Auxilia os usuários a localizar os recursos informacionais desejados, ao fornecer uma hierarquia adequada em um sistema de navegação.
- **Recuperação.** Computacionalmente, auxilia na localização de recursos informacionais em bases de dados.

Cabe ressaltar que a concepção de um vocabulário controlado pode envolver níveis de abstração e complexidade distintos quanto ao propósito de utilização e à expressividade necessária para o entendimento dos termos circunscritos ao domínio (Figura 2).

Figura 2: Complexidade no desenvolvimento de vocabulários controlados.



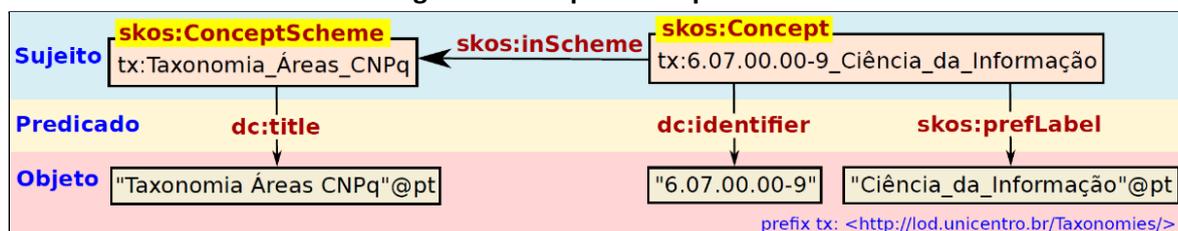
Fonte: Traduzida (NISO, 2010, p. 17).

Neste sentido, pode-se destacar quatro classes de vocabulários controlados (NISO, 2010, p. 17-18): lista de termos, anel de sinônimos, taxonomias e tesouros. Por fazer parte da base conceitual deste artigo, a seguir disserta-se com maior propriedade sobre as taxonomias.

Em suma, as taxonomias são um tipo especializado de vocabulário controlado que consiste na classificação e na ordenação hierárquica de conceitos (STEVENSON, 2006). Ou seja, ao manter a organização em uma taxonomia, para cada termo, deve-se determinar os termos relacionados e o tipo de relacionamento intratermos (de termos gerais em direção aos termos específicos ou de termos específicos no sentido aos termos gerais, por exemplo). Segundo Lambe (2007), essa organização hierárquica permite que as taxonomias determinem um esquema de classificação, mantendo uma semântica (*top-down* ou *bottom-up*) ao organizar os termos de um domínio.

Tecnicamente, como compreendido em Baeza-Yates e Ribeiro-Neto (2011), com os recentes avanços tecnológicos, organizar e representar informação e conhecimento com taxonomias passou a incorporar importantes elementos da Web Semântica. Tecnicamente, se refere ao SKOS (W3C, 2021a), o qual provê um modelo descrito em RDF (*Resource Description Framework*) para estruturação do conteúdo de vocabulários controlados, promovendo o compartilhamento do conteúdo representado na web. A Figura 3 ilustra alguns recursos RDF, ao utilizar o SKOS para representar um conteúdo abstraído a partir da Tabela de Áreas do Conhecimento do CNPq (LATTES, 2021).

Figura 3: Exemplos de triplas RDF



Fonte: Elaborada pelos autores.

Estruturalmente, uma tripla RDF é composta por três partes: sujeito, predicado e objeto, de tal forma que um sujeito é relacionado a um objeto mediante um predicado. Considerando a Figura 3, pode-se perceber uma tripla RDF em que um sujeito identificado por "tx:Taxonomia_Áreas_CNPq" tem um predicado (dc:title) que aponta ao objeto cujo conteúdo é "Taxonomia Áreas CNPq". Outro exemplo de tripla RDF utiliza o predicado

“skos:inScheme” para expressar que o sujeito “tx:6.07.00.00-9_Ciência_da_Informação” é vinculado ao sujeito/objeto “tx:Taxonomia_Áreas_CNPq”, fato que esquematicamente com SKOS representa que uma instância de “skos:Concept” pertence a um “skos:ConceptScheme”.

No tocante ao SKOS, é importante compreender que este sistema de organização de conhecimento é a recomendação do *World Wide Web Consortium* (W3C, 2021a) para interoperar vocabulários controlados na web. Neste sentido, para manter a estrutura semântica de uma taxonomia, o SKOS define alguns recursos RDF (Quadro 1).

Quadro 1: Principais recursos RDF do SKOS para modelagem de taxonomias.

Recurso	Definição
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→B), representando que A tem em B um entendimento semanticamente mais amplo. Exemplo: “tx:Representação_da_Informação” skos:broader “tx:Ciência_da_Informação”.
skos:narrower	relação hierárquica entre dois skos:Concept (A→B) e inversa de skos:broader, mapeando o entendimento mais restrito de um conceito em relação a outro. Exemplo: “tx:Ciência_da_Informação” skos:narrower “tx:Representação_da_Informação”.
skos:related	relação associativa entre dois skos:Concept (A↔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.

Fonte: (W3C, 2021a).

Para melhor entender os recursos RDF do SKOS, a subseção “4.2 – Enriquecimento” exemplifica como os referidos recursos organizam e representam os elementos das taxonomias utilizadas na pesquisa.

3 ESCOLHA E PERCURSO METODOLÓGICOS

Esta seção aborda as escolhas metodológicas, adicionalmente, enumera os conjuntos de dados utilizados e as tecnologias pertinentes à realização do estudo.

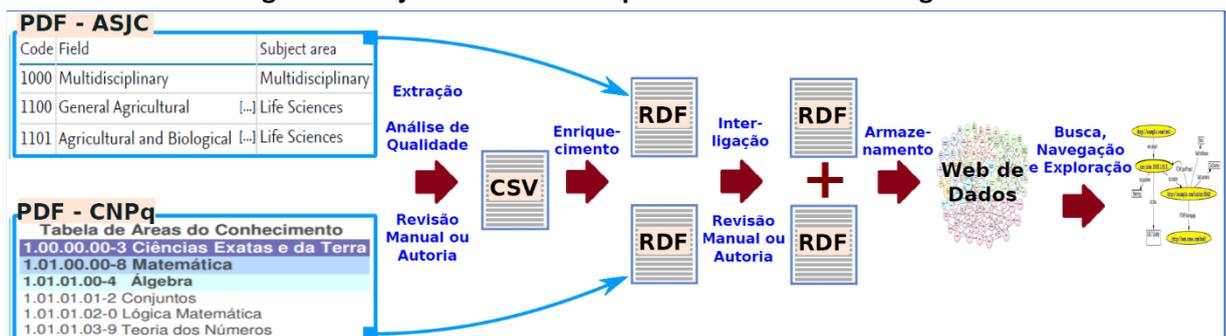
No tocante às escolhas, baseando-se em Gil (1991), a pesquisa delinea-se como: **i)** aplicada, ao utilizar insumos conceituais, procedimentais e ferramentais já estabelecidos

para implementar uma solução tecnológica; **ii)** exploratória, por aprimorar as experiências com os fatos produzidos durante o desenvolvimento do estudo; e **iii)** participante, haja vista que a pesquisa é desenvolvida no âmbito da interação dos pesquisadores quando do desenvolvimento da solução tecnológica.

Em relação aos materiais utilizados, são considerados dois conjuntos de dados: **i)** a Tabela de Áreas de Conhecimento do CNPq (LATTES, 2021); e **ii)** a Lista de Áreas Temáticas da Scopus para classificação de periódicos científicos (SCOPUS, 2021), denominada *All Scientific Journals Classification Codes (ASJC)*.

Como percurso metodológico (Figura 4), para evoluir as taxonomias em consonância aos preceitos do SKOS, são seguidas algumas atividades do ciclo de vida de dados denominado *Linked Data Lifecycle* (AUER, 2014). As atividades são:

Figura 4: Conjuntos de dados e procedimento metodológico.



Fonte: Elaborada pelos autores.

- **Exatção.** Dados estruturados ou não estruturados são capturados a partir de fontes originais e são estagiados para serem convertidos ao formato RDF.
- **Análise de Qualidade.** Aspectos de integridade, precisão, consistência, compreensibilidade e validade de dados são observados.
- **Revisão Manual ou Autoria.** Tarefas de editoração de dados são executadas para corrigir inconsistências nos dados estagiados ou produzir novos dados.
- **Enriquecimento.** Aumento da expressividade semântica é realizado, ao utilizar modelos para representar os dados como recursos RDF.
- **Interligação.** Recursos RDF são relacionados aos recursos de outras fontes, ampliando os contextos de pesquisa e consulta.
- **Armazenamento.** Recursos RDF são gerenciados mediante uso de *triplestores* (WIKIPEDIA, 2021) para potencializar a publicação e o reúso de dados na web.

- **Busca, Navegação ou Exploração.** Técnicas para recuperação são usadas para manipular ou visualizar os recursos RDF em diferentes aplicações.

Baseado nos princípios da abertura de dados, salienta-se que o procedimento metodológico subsidia a elevação dos dados das fontes originais ao nível máximo de abertura de dados (de arquivos PDF - *Portable Document Format* - 1ª estrela em direção aos recursos RDF - 5ª estrela), segundo a classificação proposta Tim Bernes-Lee (5-STAR, 2021). Isso permite que as taxonomias desenvolvidas sejam utilizadas livremente em soluções baseadas na web, além da Plataforma IAracária.

Por fim, cabe mencionar as ferramentas computacionais utilizadas no desenvolvimento da pesquisa. Neste sentido, são enumeradas:

- **Excel®.** Editor de planilhas eletrônicas da Microsoft©. Devido ao conhecimento comum dos autores, o referido editor é utilizado para organizar e revisar os conjuntos de dados, auxiliando as atividades geograficamente dispersas decorrentes da Pandemia COVID-19.
- **Open Refine®/LODRefine®.** Open Refine é uma ferramenta *open source* para trabalhar com dados desorganizados, englobando várias funções para limpeza, organização e transformação de dados (OPENREFINE, 2021). Adicionalmente, utiliza-se o pacote de extensão LODRefine, o qual incorpora as funções para evoluir conjuntos de dados para recursos RDF (GITHUB, 2021).
- **OpenLink Virtuoso®.** Sistema universal na versão *open source* para gerenciamento de dados baseados no modelo RDF na web (OPENLINK, 2021).
- **SPARQL.** Acrônimo recursivo de *Sparql Protocol and Rdf Query Language*, é uma linguagem utilizada para consultar dados descritos em triplas RDF (W3C, 2021b).

4 RESULTADO

Inspirando-se no *Linked Data Lifecycle* (AUER, 2014), a discussão do resultado alcançado é organizada em cinco subseções: **i)** extração - análise de qualidade – revisão manual ou autoria; **ii)** enriquecimento; **iii)** interligação – revisão manual ou autoria; **iv)** armazenamento; e **v)** busca, navegação e exploração.

4.1 Extração - análise de qualidade – revisão manual ou autoria

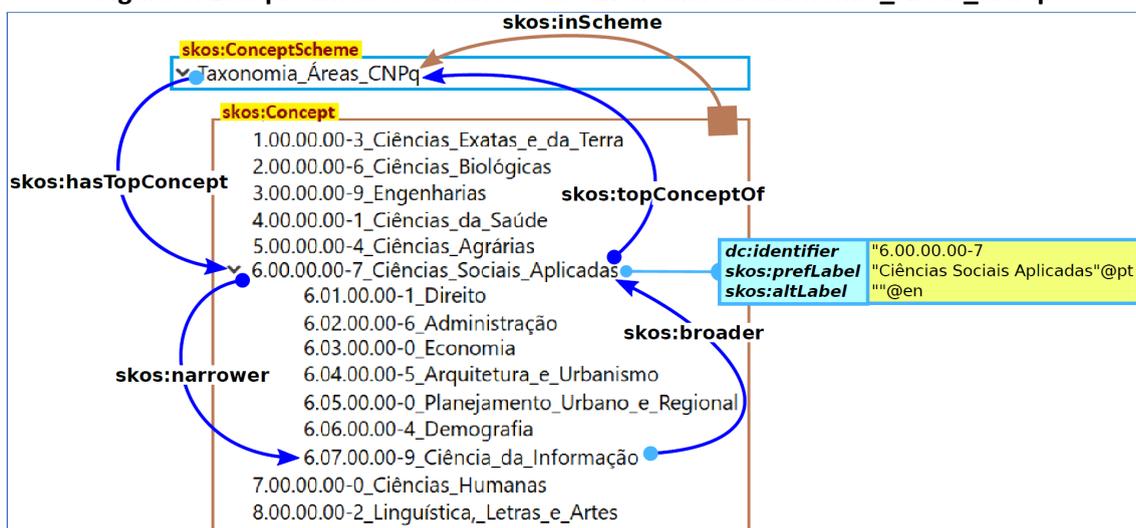
Como etapa inicial, os conjuntos de dados da Tabela de Áreas de Conhecimento do CNPq (LATTES, 2021) e a lista ASJC para classificação de periódicos científicos (SCOPUS, 2021)

foram recuperados de fontes disponíveis na web. De posse dos conjuntos de dados, na atividade de Extração foram desempenhadas as tarefas de: **i)** a migração dos dados do formato PDF ao formato XLS (*eXcel Spreadsheet*); **ii)** a limpeza, a correção e a tabulação de dados; **iii)** para o conjunto de dados ASCJ, a tradução de seus termos do idioma inglês para o idioma português; e **iv)** a conversão dos dados para o formato CSV (*Comma-Separated Value*). Nesta etapa, utilizou-se o editor de planilhas Excel®.

4.2 Enriquecimento

Nesta atividade, os dados convertidos ao formato CSV foram importados para um projeto na ferramenta OpenRefine (OPENREFINE, 2021). Com a ferramenta, mediante a utilização do pacote LODRefine® (GITHUB, 2021), os dados foram evoluídos para o formato RDF. Resumidamente, com a LODRefine®, os dados foram mapeados para o modelo RDF expresso pelo SKOS. Ou seja, para cada conjunto de dados (Tabela de Áreas de Conhecimento do CNPq e a lista ASJC), uma taxonomia é descrita no formato RDF. A seguir, considere a Figura 5, a qual exemplifica como a taxonomia da Tabela de Áreas de Conhecimento do CNPq é modelada com o SKOS.

Figura 5: Enriquecimento semântico da taxonomia tx:Taxonomia_Áreas_CNPq.



Fonte: Elaborada pelos autores.

Em suma, para cada registro do conjunto de dados da Tabela de Áreas de Conhecimento do CNPq, foi criada uma instância de `skos:Concept` (vide "6.00.00.00-7_Ciências_Sociais_Aplicadas"). A referida instância é denominada conforme a seguinte regra: concatenação código de identificação da área/subárea de conhecimento e nome da área/subárea de conhecimento, substituindo os caracteres de espaço por "_". Ainda

em relação à instância, são definidos os valores para as propriedades `dc:identifier`, `skos:prefLabel` e `skos:altLabel`, se houver.

Em seguida, a instância criada é relacionada com outras instâncias, preservando as relações hierárquicas de ancestralidade e descendência de uma taxonomia (neste caso, `skos:broader` e `skos:narrower`, respectivamente).

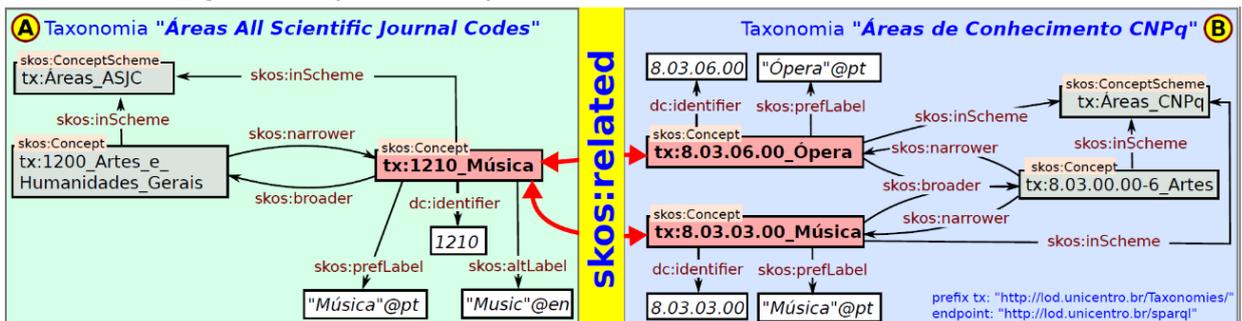
Durante a etapa de enriquecimento, também é criada uma instância de `skos:ConceptScheme` (vide no topo da Figura 5 – Taxonomia_Áreas_CNPq), a qual representa a taxonomia em desenvolvimento. Para a taxonomia criada, são agregadas todas as instâncias de `skos:Concept` mediante o uso da relação `skos:inScheme`. Por fim, são inter-relacionados todos os conceitos principais da taxonomia mediante as relações `skos:hasTopConcept` e `skos:topConceptOf`.

Salienta-se que esforços semelhantes foram empregados para desenvolver a taxonomia inerente à lista ASJC.

4.3 Interligação – revisão manual ou autoria

Nesta etapa do percurso metodológico, os recursos RDF das taxonomias criadas foram inter-relacionados de forma manual, com a relação `skos:related`.

Figura 6: Proposta de mapeamento semântico entre duas taxonomias.



Fonte: Elaborada pelos autores.

A Figura 6 exemplifica o referido processo, no qual a instância denominada "tx:1210_Música" é associada a outras duas instâncias da taxonomia "Áreas de Conhecimento do CNPq" (neste caso, "tx:8.03.03.00_Música" e "tx:8.03.06.00_Ópera").

Como ferramenta computacional auxiliar desta etapa, destaca-se o uso do editor de planilhas Excel®, pois envolveu a autoria de relacionamentos intratermos de forma manual.

4.4 Armazenamento

Na etapa de armazenamento, as taxonomias foram disponibilizadas abertamente na web, no *endpoint* <http://lod.unicentro.br/sparql>, dando origem a um grafo denominado <http://lod.unicentro.br/Taxonomies/>. Cabe ressaltar que disponibilidade do grafo na web é intermediado por uma instância do servidor OpenLink Virtuoso®.

Figura 7: Armazenamento das taxonomias – grafo <<http://lod.unicentro.br/Taxonomies/>>.



Fonte: Elaborada pelos autores.

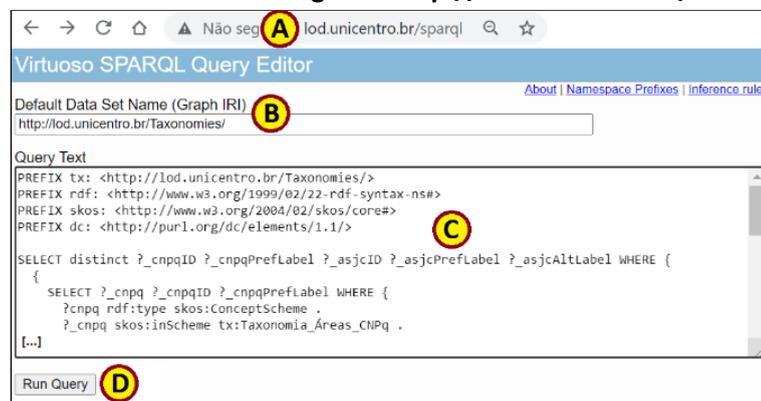
A Figura 7 representa como ocorre o armazenamento. Cabe ressaltar que a partir do referido *endpoint*, pode-se realizar consultas SPARQL, explorando os recursos disponibilizados, conforme descrito a seguir.

4.5 Busca, navegação e exploração

Para explorar o grafo <http://lod.unicentro.br/Taxonomies/> na web, tem-se duas possibilidades: i) através de *links* encurtados; e ii) por consultas SPARQL.

Com os *links* encurtados <http://tinyurl.com/hy5u7fdk> e <http://tinyurl.com/9wnztuk4>, pode-se recuperar todo o conteúdo das taxonomias denominadas “Áreas de Conhecimento CNPq” e “Áreas All Scientific Journal Codes”, respectivamente.

Figura 8: Acesso aos recursos do grafo <<http://lod.unicentro.br/Taxonomies/>>.



Fonte: Elaborada pelos autores.

Na segunda forma, com consultas SPARQL, a exploração ocorre ao executar os seguintes passos (Figura 8):

- A) acessar o *endpoint* <http://lod.unicentro.br/sparql>;
- B) definir o grafo a ser explorado, neste caso <http://lod.unicentro.br/Taxonomies/>;
- C) desenvolver uma consulta SPARQL; e
- D) executar a consulta.

Na forma de uma prova de conceito, a Listagem 1 codifica uma consulta de exploração das taxonomias compartilhadas, que ao ser submetida no *endpoint*, lista um conjunto de recursos conforme exemplificado no Quadro 2.

Listagem 1: Consulta SPARQL para recuperação do mapeamento entre as taxonomias.

```

0 PREFIX tx: <http://lod.unicentro.br/Taxonomies/>
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
0 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
2 PREFIX dc: <http://purl.org/dc/elements/1.1/>
0
3 SELECT DISTINCT ?_cnpqID ?_cnpqPrefLabel ?_asjcID ?_asjcPrefLabel ?_asjcAltLabel
0 WHERE {
4   {
0     SELECT ?_cnpq ?_cnpqID ?_cnpqPrefLabel WHERE {
5       ?_cnpq rdf:type skos:ConceptScheme .
0       ?_cnpq skos:inScheme tx:Taxonomia_Áreas_CNPq .
6       ?_cnpq rdf:type skos:Concept .
0       ?_cnpq skos:inScheme ?_cnpq .
7       ?_cnpq dc:identifier ?_cnpqID .
0       ?_cnpq skos:prefLabel ?_cnpqPrefLabel .
8     }
0   }
9 OPTIONAL
1   {
0     SELECT ?_cnpq ?_asjcID ?_asjcPrefLabel ?_asjcAltLabel WHERE {
1       ?_asjc rdf:type skos:ConceptScheme .
1       ?_asjc skos:inScheme tx:Taxonomia_All_Science_Journals_Classification_Codes .
1       ?_asjc rdf:type skos:Concept .
2       ?_asjc skos:inScheme ?_asjc .
1       ?_asjc dc:identifier ?_asjcID .
3       ?_asjc skos:prefLabel ?_asjcPrefLabel .
1       ?_asjc skos:altLabel ?_asjcAltLabel .
4       ?_asjc skos:related ?_cnpq .
1     }
5   }
1 } ORDER BY ?_cnpqID
6
7
1
8
1
9
2
2
0
2
2
3
2
4

```

2	
5	
2	
6	
2	
7	
2	
8	
2	
9	
3	
0	

Fonte: Elaborada pelos autores.

Quadro 2: Resultado parcial da consulta SPARQL da Listagem 1.

<u>cnpqID</u>	<u>cnpqPrefLabel</u>	<u>asjcID</u>	<u>asjcPrefLabel</u>	<u>asjcAltLabel</u>
[...]	[...]	[...]	[...]	[...]
"6.00.00.00-7"	"Ciências Sociais Aplicadas"@pt			
"6.07.00.00-9"	"Ciência da Informação"@pt	"3309"	"Biblioteconomia e Ciência da Informação"@pt	"Library and Information Sciences"@en
"6.07.01.00-5"	"Teoria da Informação"@pt	"3309"	"Biblioteconomia e Ciência da Informação"@pt	"Library and Information Sciences"@en
"6.07.01.01-3"	"Teoria Geral da Informação"@pt	"3309"	"Biblioteconomia e Ciência da Informação"@pt	"Library and Information Sciences"@en
"6.07.01.02-1"	"Processos da Comunicação"@pt	"3309"	"Biblioteconomia e Ciência da Informação"@pt	"Library and Information Sciences"@en
"6.07.01.03-0"	"Representação da Informação"@pt	"3309"	"Biblioteconomia e Ciência da Informação"@pt	"Library and Information Sciences"@en
[...]	[...]	[...]	[...]	[...]

Fonte: Elaborado pelos autores.

Ressalta-se que a consulta exemplo recupera a interligação entre os elementos das duas taxonomias compartilhadas, mantendo a estrutura hierárquica *top-down* da taxonomia “Áreas de Conhecimento do CNPq”, ao ordenar a consulta pelo código identificador do CNPq. Ou seja, ao observar os registros listados no Quadro 2, é possível notar que a ordem de recuperação dos recursos parte das grandes áreas de conhecimento (no caso, “6.00.00.00-7 Ciências Sociais Aplicadas”) em direção às áreas com suas respectivas subdivisões (“6.07.00.00-9 Ciência da Informação” → “6.07.01.00-5 Teoria da Informação” → “6.07.01.01-3 Teoria Geral da Informação”).

Como contribuição, com base na consulta exemplo, outras formas de recuperação podem ser inspiradas e desenvolvidas, fomentando o compartilhamento e o reuso das taxonomias em outros ambientes baseados na web.

5 DISCUSSÃO

O trabalho contribui à Ciência da Informação, principalmente, com a interligação das taxonomias “Áreas de Conhecimento CNPq” e “Áreas All Scientific Journal Codes” e seu

compartilhamento na web. Cabe ressaltar que o esforço empreendido potencializa os benefícios de reuso, compreensão, ligação, descoberta, confiança, acesso, interoperabilidade e processabilidade de recursos na web (LÓSCIO; BURLE; CALEGARI, 2021). Por exemplo, aplicações da Ciência da Informação podem capturar abertamente os recursos aderentes à categorização itens de dados de acordo com os domínios científicos descritos por termos nos idiomas português e inglês. Ainda no contexto do reuso de recursos, as taxonomias compartilhadas também podem ser empregadas para tradução, indicação de associações, rotulagem e navegação em outros sistemas de informação (NISO, 2010, p. 11).

Pontualmente, ao que se refere à Plataforma IAracária, as taxonomias desenvolvidas configuram-se como parte de uma solução sociotécnica para mediar um ambiente de cocriação e coprodução de conhecimento. A referida plataforma tem como principais usuários as Instituições de Ciência, Tecnologia e Inovação e seus colaboradores. Por meio da Plataforma IAracária, as instituições podem formar redes colaborativas de pesquisadores (os NAPIs) para promover soluções baseadas em Ciência, Tecnologia e Inovação. Diante disso, o componente “sócio” é representado pelo capital intelectual e capital social associados aos pesquisadores que possuem seu currículo cadastrado na Plataforma Lattes®. Já a componente “técnica” constitui a solução computacional que usa as taxonomias para identificação e mobilização do componente “sócio”, potencializando a cocriação e a coprodução de conhecimento. Neste sentido, cabe ressaltar que a pesquisa descrita neste artigo se insere em uma longa jornada preconizada para Plataforma IAracária (da identificação e da mobilização de capital intelectual/social em direção à coprodução de conhecimento). Quanto à identificação e à mobilização, considera-se que as taxonomias conduzem consultas mais assertivas para constituição dos NAPIs, ao evidenciar a semântica entre os diferentes domínios de conhecimento e as temáticas desejadas de pesquisa/Inovação. Já em relação à coprodução, almeja-se que as taxonomias sejam úteis a própria indexação e recuperação dos itens coproduzidos.

6 CONSIDERAÇÕES FINAIS

Este artigo disserta sobre a interligação de duas taxonomias para o domínio da Ciência da Informação. Utilizando a base teórica dos vocabulários controlados, as taxonomias denominadas “Áreas de Conhecimento CNPq” e “Áreas All Scientific Journal Codes” são desenvolvidas, compartilhadas e instanciadas a partir do Sistema de Organização de

Conhecimento conhecido como SKOS – *Simple Knowledge Organization System*. Neste sentido, como contribuição à Ciência da Informação, o esforço realizado caracteriza-se como um exemplo importante para desenvolver outras taxonomias com o SKOS e publicar as taxonomias na web para a livre utilização.

Como trabalhos futuros, almeja-se promover a Curadoria Digital das taxonomias desenvolvidas e desenvolver outros vocabulários controlados no contexto da Plataforma IAraucária.

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).

REFERÊNCIAS

5-STAR. **5-Star OPEN DATA**. Disponível em: <http://5stardata.info/en/>. Acesso em: 09 abr. 2021.

AUER, S. Introduction to lod2. *In: Linked Open Data – Creating Knowledge Out of Interlinked Data*. AUER, S.; BRYL, V.; TRAMP, C. (Ed.). **Lecture Notes in Computer Science**. Springer-Verlag, 2014.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieval: The Concepts and Technology Behind Search**. 2 ed. Harlow: Pearson Education Ltd., 2011.

FAO - FOOD AND AGRICULTURE ORGANIZATION. **AGROVOC**. Disponível em: <http://aims.fao.org/agrovoc>. Acesso em: 11 abr. 2021.

FUNDAÇÃO ARAUCÁRIA. **Nota Técnica da Fundação Araucária nº 01/2019: Novos Arranjos de Pesquisa e Inovação – NAPI**. Disponível em: http://www.fappr.pr.gov.br/sites/fundacao-araucaria/arquivos_restritos/files/documento/2020-06/nota_001_napi.pdf. Acesso em: 11 abr. 2021.

GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 1991.

GITHUB. Spartika LODRefine. Disponível em: <https://github.com/sparkica/LODRefine>. Acesso em: 09 abr. 2021.

IARAUCÁRIA. **IAraucária**. Disponível em: <https://www.iAraucaria.pr.gov.br/intelectus>. Acesso em: 09 abr. 2021.

LAMBE, P. **Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness**. Oxford: Chandos Publishing, 2007.

LATTES. **Tabela de Áreas do Conhecimento**. Disponível em: <http://lattes.cnpq.br/documents/11871/24930/TabeladeAreasdoConhecimento.pdf/d192ff6b-3e0a-4074-a74d-c280521bd5f7>. Acesso em: 09 abr. 2021.

LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. **Data on the Web Best Practices - W3C Recommendation 31 January 2017**. Disponível em: <https://www.w3.org/TR/dwbp>. Acesso em: 09 abr. 2021.

NISO - NATIONAL INFORMATION STANDARDS ORGANIZATION. **ANSI/NISO Z39.19-2005 (R2010): guidelines for the construction, format, and management of monolingual thesauri**. American National Standards Institute, 2010.

OPENLINK VIRTUOSO. Disponível em: <http://virtuoso.openlinksw.com/>. Acesso em: 09 abr. 2021.

OPENREFINE. **OpenRefine: a free, open source, powerful tool for working with messy data**. Disponível em: <https://openrefine.org/>. Acesso em: 09 abr. 2021.

PELLIZZON, R. de F. Research in the area of health: 1 - DeCS database (Health Sciences Descriptors). **Acta Cirúrgica Brasileira**, v. 19, n. 2, p. 153-163, 2004. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-86502004000200013&lng=en&nrm=iso. Acesso em: 11 Abr. 2021. <https://doi.org/10.1590/S0102-86502004000200013>.

SCOPUS. **What is the complete list of Scopus Subject Areas and All Science Journal Classification Codes (ASJC)?** Disponível em: https://service.elsevier.com/app/answers/detail/a_id/15181/supporthub/scopus. Acesso em: 09 abr. 2021.

STEVENSON, J. **Dictionary of Information and Library Management**. 2 ed. Londres: A. & C. Black, 2006.

W3C – WORLD WIDE WEB CONSORTIUM. **SKOS Simple Knowledge Organization System Reference**. Disponível em: <https://www.w3.org/TR/2009/REC-skos-reference-20090818>. Acesso em: 09 abr. 2021a.

_____. **SPARQL 1.1 Query Language: W3C Recommendation**. Disponível em: <https://www.w3.org/TR/2013/REC-sparql11-query-20130321>. Acesso em: 10 abr. 2021b.

WIKIPEDIA. **Triplestore**. Disponível em: <https://en.wikipedia.org/wiki/Triplestore>. Acesso em: 10 abr. 2021.