Explorando a reconciliação de dados culturais na Wikidata: experimento aplicado com o acervo museológico do Museu Histórico Nacional

Luis Felipe Rosa de Oliveira
Doutorando em Ciência da Informação pela Universidade de Brasília (UnB) - Brasília, DF - Brasil. Mestre em Comunicação pela Universidade Federal de Goiás (UFG) – GO - Brasil.
http://lattes.cnpq.br/6498992926514286
E-mail:luisfelipeprf@gmail.com

Dalton Lopes Martins
Pós-Doutorado pela Universidade de São Paulo (USP) – SP - Brasil. Doutor em Ciência da Informação pela Universidade de São Paulo (USP) - São Paulo, SP - Brasil. Professor da Universidade de Brasília (UnB) - Brasília, DF - Brasil.
http://lattes.cnpq.br/3774617443225038
E-mail:dmartins@gmail.com

Submetido em: 22/09/2020. Aprovado em: 25/11/2020. Publicado em: 28/07/2021.

RESUMO

Este estudo foi desenvolvido sob a perspectiva da web semântica e dos dados abertos ligados, com enfoque na técnica de reconciliação de dados e busca entender como ocorre o processo de reconciliação de dados culturais com a Wikidata, através da programação de scripts na linguagem Python, com o objetivo de contribuir para o entendimento de como se dá a aplicação de uma técnica de enriquecimento semântico em bases de dados culturais. Como metodologia, são descritas as etapas de desenvolvimento dos scripts de reconciliação de dados. E, como resultados, são apresentados os produtos da aplicação dos scripts na reconciliação de parte dos dados do acervo museológico do Museu Histórico Nacional com os objetos digitais da Wikidata. Chega-se à conclusão de que o processo de descrição do desenvolvimento dos scripts permitiu compreender melhor como ocorre a reconciliação de dados em acervos culturais, de que se deve dar mais atenção à normalização dos dados do acervo, e de que esse tipo de aplicação amplia o potencial de socialização do conhecimento em rede.

Palavras-chave: Web semântica. Dados abertos ligados. Acervos culturais. Enriquecimento semântico.

Exploring the reconciliation of cultural data on wikidata: experiment applied with the museum collection of the national historical museum

ABSTRACT

This study was developed from the perspective of the semantic web and the linked open data, focusing on the data reconciliation technique. It seeks to understand how the process of reconciling cultural data with Wikidata *occurs through the programming of scripts in the Python language, with the aim of contributing to the understanding of how to apply a technique of semantic enrichment in cultural databases. As a methodology, the stages of the data reconciliation scripts development are described. And as results, the products of the scripts application are presented in the reconciliation of part of the data of National Historical Museum museological collection with the digital objects of* Wikidata. *It is concluded that the process of describing the scripts development allowed to better understand how data reconciliation occurs in cultural collections, and that more attention should be paid to the normalization of the collection data, and that this type of application expands the potential for networked socialization of knowledge.*

Keywords*: Semantic web. Linked open data. Cultural collections. Semantic enrichment.*

Explorando la reconciliación de datos culturales en wikidata: experimento aplicado con la colección del museo del museo histórico nacional

RESUMEN

Este estudio se desarrolló desde la perspectiva de la web semántica y los datos abiertos vinculados, centrándose en la técnica de conciliación de datos. Busca comprender cómo ocurre el proceso de conciliación de datos culturales con Wikidata *a través de la programación de scripts en el lenguaje Python, para contribuir a la comprensión de la aplicación de una técnica de enriquecimiento semántico en las bases de datos culturales. Como metodología, se describen las etapas de desarrollo de los scripts de reconciliación de datos. Y como resultado, los productos de la aplicación de los scripts se presentan en la conciliación de parte de los datos de la colección museológica del Museo Histórico Nacional con los objetos digitales de* Wikidata. *Se concluye que el proceso de describir el desarrollo de los scripts nos permitió comprender mejor cómo se produce la conciliación de datos en las colecciones culturales, y que se debe prestar más atención a la normalización de los datos recopilados, y que este tipo de aplicación expande un potencial de socialización del conocimiento en red.*

Palabras clave*: Web Semántica. Datos abiertos vinculados. Colecciones culturales. Enriquecimiento semántico.*

INTRODUÇÃO

A apropriação crescente dos conceitos da web semântica pelos estudos em Ciência da Informação é notável, uma temática que consolida estudos importantes da área, como construção e aplicação de ontologias, estruturas de dados semânticos e interoperabilidade de bases de dados, contextualizando-os no ambiente digital da web, agregando as condições do fenômeno de interação em rede e a alta demanda pela capacidade de lidar com o volume e a variabilidade de dados gerados a cada segundo.

Se aventurar pelo universo de possibilidades da web semântica é tentador a qualquer profissional da informação, porém esse processo pode ser um pouco frustrante em um primeiro momento. Entender quais os aspectos da web semântica podem beneficiar suas atividades é um desafio, dado que é necessário convergir diferentes técnicas e estruturas agregadas em um ambiente dinâmico de conexão e reúso de informações para a criação de produtos e serviços informacionais. É, a partir desse sentimento, que este estudo se origina da necessidade de entender como processos inerentes da web semântica se dão, e como eles podem beneficiar o acesso e reúso da informação.

Vale elencar aqui o estudo desenvolvido por Santarém Segundo (2014), que apresentou uma exploração do uso do protocolo SPARQL (protocolo para consultas e manipulação de dados estruturados semanticamente) para recuperação da informação em bases semânticas, resultando na indicação da viabilidade do protocolo para acesso a esse tipo de conteúdo semântico na web, fundamentando a aplicação efetivada no estudo aqui proposto sob o contexto das informações de cunho cultural na web.

Mais especificamente, como acervos digitais de instituições de patrimônio cultural podem incorporar os benefícios e serem enriquecidos com os produtos de aplicações semânticas. O foco da pesquisa é demonstrar meios práticos e operacionais de se valer de bases de conhecimento já existentes para efetivar a ligação de dados e ampliar a possibilidade de adoção desse recurso para ampliar e enriquecer as fontes de informação culturais.

O movimento dessas instituições de patrimônio cultural de digitalização dos objetos culturais e a disponibilização online de seus acervos entre outras consequências condicionam uma dinâmica de difusão cultural em rede, com um grande potencial de consumo e agregação com outros provedores de dados. Enxergando que tópicos da web semântica podem se valer desse potencial, e que os profissionais atuais da Ciência da Informação, com conhecimento técnico prévio, podem desenvolver recursos. Nessa perspectiva, este estudo busca entender como se dá o processo de reconciliação de dados culturais com a Wikidata através da programação de scripts na linguagem Python.

Como objeto prático do estudo, foi utilizada a base de dados do acervo museológico disponível no repositório digital do Museu Histórico Nacional (MHN) disponível para acesso no link http://mhn.acervos.museus.gov.br/reserva-tecnica/. Este acervo faz parte de um conjunto de outros acervos publicados atualmente na web, ressaltando, aqui, a iniciativa do Instituto Brasileiro de Museus em promover a disponibilização on-line dos acervos digitais dos museus sob sua gestão através do software Tainacan[1]. Esse tipo de iniciativa reforça a estruturação de um conjunto de objetos digitais culturais disponíveis na web, abrindo espaço para o alto potencial do uso de aplicações semânticas para enriquecimento de acervos na área cultural.

Desse modo, o estudo aqui proposto se estrutura sob o desenvolvimento de scripts de reconciliação de dados com o objetivo de reconciliar parte do conjunto de dados do acervo museológico do MHN com objetos digitais da Wikidata, colocando esse acervo no universo dos dados ligados, e abrindo portas para o enriquecimento de seus dados.

Este experimento foi realizado com objetivo principal de contribuir para o entendimento de como se dá a aplicação de uma técnica de enriquecimento semântico em bases de dados culturais.

[1] Tainacan - https://tainacan.org/ Acesso em 19/09/2020.

Vale ressaltar ainda que este resultado é parte de um esforço maior de pesquisa e faz parte de um projeto de doutorado em andamento, que tem por objetivo implementar um serviço de reconciliação de dados para todos os repositórios digitais disponibilizados na plataforma Tainacan no âmbito do Instituto Brasileiro de Museus.

O artigo apresenta a seguir uma contextualização dos conceitos que fundamentam a aplicação do estudo, a metodologia com a descrição do desenvolvimento dos scripts de reconciliação de dados, e, em seguida, **a análise dos resultados aponta sínteses dos produtos dos scripts e alguma**s reflexões sobre os resultados. Por fim, nas considerações finais, **são discutidos os principais pontos alcançados e** as propostas de pesquisas futuras para continuar a contribuição do entendimento da aplicação deste tipo de técnica de enriquecimento semântico em acervos digitais culturais.

WEB SEMÂNTICA E RECONCILIAÇÃO DE DADOS

A reconciliação de dados é o conceito que fundamenta a pesquisa aplicada neste estudo, porém, antes de apresentar sua definição, é importante contextualizar os conceitos de web semântica e dados abertos ligados que fundamentam a temática sob a qual este artigo foi desenvolvido.

Idealizada por Tim Berners-Lee, a web semântica é uma "extensão da web", que possibilita um fluxo de acesso à informação mais estruturado para a leitura por softwares, potencializando funcionalidades de busca e reúso da informação, o que, consequentemente, afeta a forma como os usuários acessam a informação na web, amplificando as formas de como o conteúdo pode ser gerado e consumido em rede (BERNERS-LEE; HENDLER; LASSILA, 2001).

Essa visão da web semântica na prática é mais perceptível do ponto de vista do desenvolvimento de tecnologias de acesso e reúso da informação, pois a estruturação dos dados de forma semântica na web permitirá que repositórios digitais relacionem informações de diversos provedores digitais, como por exemplo, um acervo museológico publicado através do Tainacan na web poder ter seus dados de autoria relacionados com os dados de um provedor de controle autoridades como o VIAF[2], ou ainda dados de localização com um provedor de dados geográficos como o GeoNames[3].

Os resultados desse tipo de relacionamento dos dados são muitos. Da perspectiva da qualidade da informação, esse tipo de aplicação reduz a incerteza sobre os dados, pois ao conectá-los com um provedor com controle semântico, os dados são contextualizados e recebem identificadores únicos na web. Outra perspectiva é dada a partir da geração automática de conteúdo, uma vez que os dados estão estruturados de forma semântica, é possível gerar conteúdos de maneira automática, como são feitos os painéis de informação do Google que aparecem ao se pesquisar por uma entidade importante, por exemplo, quando se pesquisa por Tim Berners-Lee no Google[4], aparece em destaque um conjunto de fotos dele, um breve resumo biográfico obtido da Wikipedia , os livros publicados por ele e algumas pesquisas relacionadas.

Para que essas possibilidades semânticas existam de maneira mais frequente e acessível, é necessário que alguns padrões, que são indicados principalmente pelo W3C (*World Wide Web Consortium*), sejam adotados e aplicados, como o RDF (*Resource Description Framework*), que permite descrever objetos digitais a partir de uma estrutura semântica de triplas, envolvendo um sujeito, um predicado e um objeto. Vale ressaltar ainda o papel importante das ontologias e dos padrões de metadados, que, quando aplicados efetivamente, abrem as portas para a estruturação semântica de bases de dados na web.

[2] VIAF - http://viaf.org/

[3] GeoNames - https://www.geonames.org/

[4] Painel de Informação do Google de Tim Berners-Lee - https://g.co/kgs/nb12Dg

Dessa forma, o papel do cientista da informação se revela imprescindível no contexto na web semântica, produzindo e apoiando estudos e projetos sobre "motores de busca, interfaces dos sistemas de informação, vocabulários controlados, indexação automática, gestão do conhecimento e inteligência competitiva" como elencado por Souza e Alvarenga (2004, p. 139-140).

E é, nessa perspectiva, que este artigo é situado, no contexto em que a Ciência da Informação apoia as instituições de patrimônio cultural na publicação de seus acervos digitais em rede, nesse caso em específico, propondo um estudo aplicado sobre o relacionamento dos dados dos acervos com provedores de dados semânticos.

E essa condição de relacionamento dos dados é posicionada sob o conceito de dados abertos ligados, que também é difundido por Berners-Lee (2006), e, em suma, é fundamentado sob quatro princípios: *Usar URIs como nomes para coisas; Usar URIs em HTTP para que as pessoas consigam acessar esses nomes; Prover informações úteis junto à URI, utilizando padrões semânticos, como RDF; Referenciar outras URIs, para que seja possível descobrir outras coisas.*

Esses princípios expressam de maneira objetiva e técnica que efetivar a publicação de dados abertos ligados envolve um processo de identificação dos objetos, como colocado acima, através de URIs (*Uniform Resource Identifier*) que são identificadores únicos na web, e certificar-se de que esses identificadores estejam acessíveis e disponíveis para serem ligados uns com os outros entre os provedores de dados da internet.

Essa estruturação de objetos na web é proveniente do contexto da web semântica, e, inclusive, sistematiza uma prática de sociabilidade em rede de informações entre as bases de conhecimento existentes, direcionando a produção e consumo de conteúdo na internet a uma maior versatilidade da relação entre o usuário e a máquina.

Dessa forma, imaginando promover essa sociabilidade em rede, algumas técnicas podem ser aplicadas para promover a contextualização semântica de bases de dados na web, uma delas é a reconciliação de dados, que, no contexto da web semântica, pode ser entendida como um dos processos inerentes ao enriquecimento semântico, que, por sua vez, constitui um conjunto de técnicas que objetivam ligar dados com bases de conhecimentos digitais (SANDERSON, 2016).

Essas bases de conhecimentos digitais podem ser compreendidas como sistemas de organização do conhecimento (KOS) disponíveis na web, e podem ser expressos tanto na forma de vocabulários controlados, quanto na forma de um repositório de objetos digitais estruturados semanticamente, como a Wikidata ou o *Geonames*, sendo chamados, nesse caso, de KOS-LOD, pois são sistemas de organização do conhecimento presentes na nuvem de dados abertos ligados (ZENG, 2019).

Uma boa forma de entender como ocorre o processo de enriquecimento semântico é refletir sobre o Framework de Enriquecimento Semântico proposto pela Europeana:

> *Análise*: a fase de pré-enriquecimento concentra-se na análise dos metadados originais, na seleção dos sistemas de conhecimento a serem ligados, e na proposição de regras para correspondência e vínculo dos metadados originais ao recurso textual disponíveis nos KOS selecionados; *Vinculação*: o processo de combinação automática entre os valores dos metadados com os valores dos objetos nos KOS, e a adição de relação contextuais entre os valores; *Acréscimo*: o processo de seleção dos valores do KOS a serem adicionados ao conjunto de dados original. Isso talvez não inclua somente conceitos em diferentes línguas, mas também conceitos mais específicos ou abrangentes (ISAAC *et al.*, 2015, p. 9).

Em síntese, o processo de enriquecimento semântico compreende um momento de análise dos metadados atuais do acervo/conjunto de dados a ser enriquecido, em que também é realizada a seleção de quais bases de conhecimento serão utilizadas como referência para a ligação dos dados e quais as especificações e regras para a efetivação da ligação dos dados.

Após a análise, o momento de vinculação se refere ao uso de técnicas automáticas/semiautomáticas de reconhecimento dos valores textuais provenientes dos metadados originais, nas bases de conhecimento on-line, bem como à descrição dos tipos de relacionamento das combinações resultantes das ligações. Por último, o acréscimo diz sobre a adição de informações presentes na base de conhecimento ao conjunto de dados enriquecido, com seus devidos relacionamentos expressos.

Esse processo que envolve o enriquecimento semântico, bem como a reconciliação de dados, é diretamente relacionado com a proposta dos dados abertos ligados, e é componente importante da estruturação e potencial aplicação em serviços futuros da web semântica entre as instituições de patrimônio cultural. Vale dizer que isso é ainda praticamente inexplorado nos repositórios digitais e serviços informacionais culturais no país, representando um ponto importante de desenvolvimento futuro para a área. É o processo que auxilia na mudança de concepção do modelo conceitual dos dados, saindo do foco no documento para o foco nas entidades. A partir da aplicação desse processo, o aprimoramento do acesso e reúso do acervo/conjunto de dados pode ser efetivado. Como consequência, pode-se obter novos serviços e produtos, gerando camadas de inovação a partir dos dados e ampliação do valor social dos acervos.

O estudo aplicado neste artigo tem o foco especificamente no tópico da reconciliação de dados, expressa na etapa de vinculação do enriquecimento semântico em que ocorre "o processo de combinação automática entre os valores dos metadados com os valores dos objetos nos KOS" (ISAAC *et al.*, 2015, p. 9), e, ainda no caso deste estudo, limita-se a implementar esse processo, sem, posteriormente, efetuar adição de relações contextuais.

O sistema de organização do conhecimento KOS-LOD utilizado neste experimento foi a Wikidata, parte do universo informacional da Wikipedia e que compartilha das mesmas características de colaboração em rede.

A Wikidata originalmente foi pensada para criar objetos digitais e relacioná-los ao conteúdo de páginas da Wikipedia , estruturada sob uma ótica semântica (<objeto><relação><objeto>). No entanto, a Wikidata logo ganhou independência e se tornou referência como base de conhecimento digital pela grande quantidade de informação estruturada e relacionada (VRANDEČIĆ; KRÖTZSCH, 2014). Atualmente, a Wikidata conta com mais de 71 milhões[5] de objetos digitais interligados em diversas linguagens.

METODOLOGIA

O método aplicado neste estudo se limita à forma de como foram concebidos os scripts de reconciliação de dados, já que estes foram desenvolvidos justamente para atender à expectativa do estudo de permitir entender melhor o processo de reconciliação semiautomática de dados, utilizando a linguagem de programação Python.

A operação de reconciliação de dados foi composta por dois scripts, sendo o primeiro deles, identificação de instâncias[6], com a função de identificar instâncias (classes de objetos) da Wikidata que melhor representam os valores de cada metadado da base de origem, uma vez sinalizada qual a instância. O segundo script de reconciliação de dados[7] realiza o processo de busca por objetos da Wikidata que representem os valores procurados, filtrando cada valor pela sua respectiva instância sinalizada no primeiro script, perfazendo, assim, um caminho de ligação de dados entre uma base de dados e um sistema de organização do conhecimento.

[5] Wikidata Statistics - https://www.Wikidata.org/wiki/Wikidata:Statistics/pt-br

[6] Script de identificação de instâncias - https://github.com/luisfeliperd/Wikidata_sparql/blob/master/Wikidata_metadata.py

[7] Script de reconciliação de dados – https://github.com/luisfeliperd/Wikidata_sparql/blob/master/Wikidata_value.py

O ambiente computacional utilizado para desenvolver os scripts foi constituído por um notebook com acesso à internet banda larga cabeada de 35Mb, cujo hardware é composto por um processador de 4 núcleos físicos de até 3.80GHz de frequência, 16GB de memória RAM, utilizando como unidade de armazenamento um SSD e como sistema operacional o Windows 10. Ressaltando ainda que não foi utilizado nenhum servidor de banco de dados, os dados coletados foram armazenados no formato CSV e processados posteriormente para análise em planilhas.

A título de experimentação, os dados utilizados nos scripts foram obtidos exportando a coleção de acervo museológico do Museu Histórico Nacional no formato CSV[8]. Essa base de dados é formada por 774 itens, e foram escolhidos três metadados para o processo de reconciliação, são eles: autor, técnica e material. A tabela 1 abaixo apresenta a quantidade de valores para cada metadado, sendo que essa quantidade não expressa o total de itens da base devido à existência de valores vazios, além disso, como os valores se repetem, a coluna *valores distintos* apresenta a quantidade de valores únicos sem repetição.

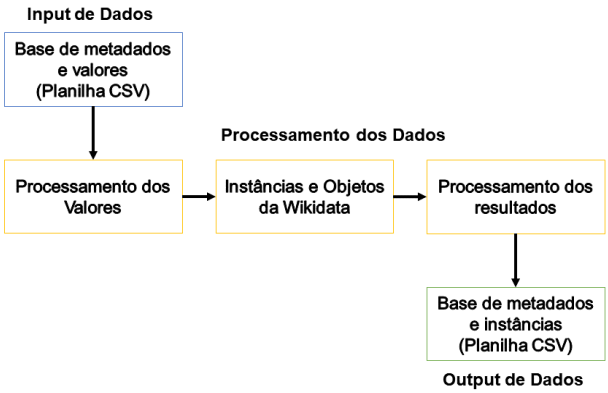
Tabela 1 – Síntese da quantidade de valores por metadado. Valores sintetizados a partir dos dados do acervo museológico do Museu Histórico Nacional

Metadado	Valores	Valores Distintos
Autor	646	241
Técnica	710	43
Material	775	51

Fonte: Dados da pesquisa (2020).

De forma geral, ambos os scripts funcionam a partir de um fluxo semelhante de processamento dos dados, diferenciado apenas nas funcionalidades aplicadas, como apresenta o fluxograma da figura 1.

Figura 1 – Fluxo básico dos scripts de reconciliação de dados



Fonte: Dados da pesquisa (2020).

O *input de dados* dos scripts ocorre através de uma planilha no formato CSV com a base de dados a ser reconciliada, no caso, no script de reconciliação de dados, a planilha de seleção de instâncias (resultante do script de identificação de instâncias) também é utilizada como entrada para definir em qual instância buscar de cada conjunto de valores.

O *processamento de dados* dos scripts perfaz o caminho de procurar por cada valor da base de dados de origem (acervo do MHN no caso) na Wikidata através de uma consulta SPARQL[9] no *endpoint* da API de consultas[10], que busca pelos valores nos rótulos de objetos na Wikidata. Essas consultas foram formadas, utilizando o tutorial produzido pela própria Wikidata[11], e o link de consulta SPARQL[12] para elaborar testes. No caso do script de identificação de instâncias, os resultados são focados nas possibilidades de identificar que tipo de instância seria sugerido para cada valor e, no caso do script de reconciliação de dados, os resultados obtidos de cada instância que expressam as possibilidades de objetos da Wikidata que referenciem o valor procurado.

[8] Os dados do acervo digital do museu podem ser acessados na função "Ver como..." da interface do Tainacan do MHN, por esse ponto de acesso, é possível acessar a base via API, uma planilha HTML e uma planilha CSV.

[9] O que são consultas SPARQL? - https://www.w3.org/TR/rdf-sparql-query/

[10] *Endpoint* da API de consulta SPARQL - https://query.Wikidata.org/sparql

[11] Tutorial SPARQL Wikidata - https://www.Wikidata.org/wiki/Wikidata:SPARQL_tutorial

[12] Link para consultas SPARQL na Wikidata - https://query.Wikidata.org/

Vale reforçar que instância aqui é considerada como um conceito abstrato ou uma entidade, por exemplo, Autor, e que o resultado de cada instância é, por exemplo, um autor específico ou um caso de uma instância, como Machado de Assis. Estruturar a busca por reconciliação dessa maneira amplia a possibilidade de filtro de uma base semântica, dado que primeiro se identifica a instância e depois se identificam os casos de uma instância que podem se conectar.

Já o *output de dados*, consiste no processamento dos resultados da consulta SPARQL, que são obtidos no formato JSON (Notação de Objetos JavaScript). No caso do script de identificação de instâncias, os dados de *nome da instância* e *QID da instância* (identificador dos objetos na Wikidata) são obtidos para cada valor procurado, e, para cada metadado, é mensurada a ocorrência das instâncias em seus valores, o que permite criar uma planilha com as instâncias que mais ocorreram para cada metadado, permitindo ao usuário escolher entre as indicações de qual instância melhor se adequa para representar os metadados da base de dados de origem. No caso do script de reconciliação de dados, cada valor da base de dados é buscado na Wikidata, de acordo com a instância selecionada para cada metadado na planilha resultante do script anterior, os termos dos valores são comparados com os termos dos rótulos dos objetos da Wikidata, e, de acordo com a semelhança entre os termos, uma pontuação é gerada, indicando qual objeto potencialmente representa o valor procurado. O método tem um princípio estatístico, descrito abaixo pela biblioteca *fuzzywuzzy*[13] , apresentando qual a maior probabilidade de um conjunto de objetos informacionais ser de uma determinada instância pela recorrência e proximidade de vários termos relacionados a ela, além de, dada uma instância, qual caso melhor se ajusta a um valor de ocorrência para um metadado específico, indicando o caso mais próximo de ocorrência.

A seguir, o desenvolvimento de cada script será descrito com base em 5 etapas: descrição das bibliotecas utilizadas, do processo de leitura dos dados da base de origem, da consulta dos valores na Wikidata e do processamento dos resultados obtidos.

Quanto ao script de identificação de instâncias, as bibliotecas utilizadas foram a *requests*[14] para fazer a consulta dos valores ao *endpoint* da API de consulta SPARQL da Wikidata, também foi utilizada a biblioteca *pandas*[15] para lidar com a estruturação dos dados no script, como leitura da base de dados, armazenamento dos dados coletados e exportação dos resultados, além dessas também foi utilizada a biblioteca *time*[16] para parar o script por alguns segundos a cada consulta à Wikidata, evitando um possível bloqueio de segurança da API, por fim, foi utilizada a biblioteca *datetime*[17] para calcular o tempo gasto para consultar os valores.

A leitura de dados para o script de identificação de instâncias ocorreu primeiramente lendo os registros da planilha da base de dados de experimentação do MHN, cuja composição foi mencionada na tabela 1, acima. Para cada metadado, foi recuperado cada valor e para cada valor realizada uma consulta. Não foi realizado nenhum processo de normalização dos dados, dessa forma, termos sem padronização, e valores em branco ocorrem nessa base de dados.

Outro ponto a ser ressaltado é que como cada objeto pode ter valores de metadados iguais, por exemplo, mesmos autores, ou mesma técnica de produção, os valores na base se repetem. Dessa forma, foi aplicada uma verificação no script que pula a consulta de um valor se ele já foi verificado anteriormente, isso reduz a demanda à API e otimiza a síntese dos resultados.

[13] Biblioteca *fuzzywuzzy* - https://pypi.org/project/fuzzywuzzy/

[14] Biblioteca *requests* - https://pypi.org/project/requests/#description

[15] Biblioteca *pandas* - https://pandas.pydata.org/docs/

[16] Biblitoeca *time* - https://docs.python.org/3/library/time.html

[17] Biblioteca datetime https://docs.python.org/3/library/datetime.html

Ainda foi necessário tratar a existência de múltiplos valores nos campos de técnica e material, por exemplo, alguns itens têm como material óleo e tela, o que ocorre na planilha com a notação "óleo||tela", com os termos divididos por *dual pipe* ("||"), desse modo, foi preciso definir um processo de separação desses valores e consulta a cada um individualmente.

Já na etapa de consulta dos valores na Wikidata, cada valor, após passar pelo processo de leitura de dados citado no parágrafo anterior, foi inserido em uma consulta SPARQL (figura 2) embutida no código. Essa consulta SPARQL foi estruturada para recuperar o identificador do objeto da Wikidata em "?sujeito", o identificador da instância do objeto encontrado em "?instancia_de_que" e o rótulo da instância em "?instancia_de_queLabel", a consulta é feita de forma unificada nos idiomas português de Portugal, português do Brasil e inglês, de forma que são consultados objetos da Wikidata que tenham o rótulo igual ao valor consultado. A consulta ainda prevê que não sejam recuperadas instâncias referentes à "categoria da Wikimedia" (Q4167836) e "páginas de desambiguação da Wikimedia" (Q4167410).

A cada consulta foi aplicada uma espera de 3 segundos visando a evitar problemas de bloqueio de segurança da API, além disso, o resultado da consulta pode retornar um JSON vazio se nenhuma instância for identificada, questão que foi tratada no script, pulando os valores que não retornaram resultados. Dessa forma, uma consulta válida retorna um JSON com dados sobre o nome da instância e o identificador da instância na Wikidata.

Figura 2 – Consulta SPARQL do script de identificação de instâncias

```
SELECT DISTINCT ?sujeito ?instancia_de_que ?instancia_de_queLabel WHERE {

                        { ?sujeito ?label "%s". }
                        UNION
                        { ?sujeito ?label "%s"@en. }
                        UNION
                        { ?sujeito ?label "%s"@pt-br. }

                        ?sujeito wdt:P31 ?instancia_de_que.

                        FILTER(?instancia_de_que NOT IN(wd:Q4167836, wd:Q4167410))

                        SERVICE wikibase:label { bd:serviceParam wikibase:language "pt-br", "pt", "en". }
}
```

Fonte: Dados da Pesquisa (2020).

Por fim, o processamento dos dados coletados foi executado em duas etapas, a primeira foi o armazenamento dos resultados de cada consulta SPARQL, juntamente com as informações dos valores consultados, o que consistiu em um dataframe (estrutura de dados de múltiplas variáveis no *pandas*) com o nome do metadado, o nome da instância e o identificador da instância recuperada. Após a coleta dos dados para todos os metadados, um novo dataframe foi construído, sintetizando os dados, foi calculada a ocorrência dos identificadores das instâncias, e consideradas apenas as 5 instâncias que mais se repetiram para cada metadado no esforço de identificar as instâncias mais representativas.

Uma planilha com essas 5 instâncias e suas respectivas ocorrências para cada metadado é o resultado deste script, essa mesma planilha contém uma coluna denominada "best_option" para que o usuário sinalize com um "x" qual das instâncias seria a mais representativa para cada metadado e, assim, utilizar essa planilha como *input* do próximo script, apontado em qual instância procurar os valores da base de dados para cada metadado.

Já quanto ao script de reconciliação de dados, foram utilizadas as mesmas bibliotecas *requests*, *pandas*, *time e datetime* utilizadas no script anterior, com a adição da biblioteca *collections* para usar especificamente a função *defaultdict*, que permite a criação de dicionários de listas para auxiliar na estruturação dos dados, e também a biblioteca *fuzzywuzzy* para calcular uma pontuação de semelhança de termos entre os valores da base de dados e os rótulos de objetos da Wikidata.

A leitura dos dados para esse script ocorreu da mesma forma como no script anterior, lendo a base de dados e verificando cada valor para cada metadado, atentando-se aos mesmos princípios de valores repetidos e a valores múltiplos. A diferença principal foi a leitura da planilha resultante do script de identificação de instâncias com a informação de qual instância usar para procurar os valores em cada metadado. A partir dela, o dado do identificador da instância foi adicionado à consulta SPARQL e permitiu pesquisar os valores em uma instância específica.

A etapa de consulta dos valores no script de reconciliação de dados consistiu na verificação dos valores de cada metadado da base de dados na Wikidata via consulta SPARQL (figura 3), que retornava como dados, o identificador do objeto na Wikidata ("?sujeito"), o rótulo principal do objeto ("?sujeitoLabel") e os rótulos alternativos do objeto ("?sujeitoAltLabel"). Essa consulta busca os valores da base de dados que tenham o termo semelhante aos rótulos dos objetos da Wikidata. Além disso, o dado da instância também é utilizado na penúltima linha da consulta para refinar a consulta dos valores a objetos da instância referente ao metadado.

Figura 3 – Consulta SPARQL do script de reconciliação de dados.

```
SELECT DISTINCT ?sujeito ?sujeitoLabel ?sujeitoAltLabel WHERE {

                                { ?sujeito ?label "%s". }
                                UNION
                                { ?sujeito ?label "%s"@en. }
                                UNION
                                { ?sujeito ?label "%s"@pt-br. }

                                ?sujeito wdt:P31 wd:%s

                                SERVICE wikibase:label { bd:serviceParam wikibase:language "pt-br", "pt", "en". }
}
```

Fonte: Dados da Pesquisa (2020).

Devido ao volume de consultas, foi preciso adicionar um *loop* ao script que verifica se a consulta à API foi bem sucedida, caso contrário é adicionada uma espera de 5 minutos até a realização da próxima consulta, visando a contornar um possível bloqueio momentâneo da API. Esse processo é repetido pelo menos 5 vezes, até que o acesso à API seja estabelecido ou, então, o script retorna ao erro de acesso.

O processamento dos resultados, no caso desse script, executa uma etapa de armazenamento dos identificadores dos objetos encontrados e dos seus respectivos rótulos, sendo que para cada rótulo um cálculo de semelhança entre o termo do valor consultado e os rótulos dos possíveis objetos da Wikidata encontrados é efetuado. Este cálculo é executado usando a biblioteca *fuzzywuzzy* que utiliza o princípio do cálculo da distância Levenshtein, cujo princípio "é definir a distância entre duas palavras com base no número de operações necessárias para torná-las iguais" (RUBERTO; RODRIGO, 2017, p. 31), essa operação no script retorna uma pontuação de semelhança entre termos no intervalo de 0 para nenhuma semelhança e 100 para termos iguais. Dessa forma, a segunda etapa do processamento de dados calcula uma média dos scores para cada objeto da Wikidata encontrado, apontando qual deles tem uma semelhança de termos maior com o valor consultado.

Por fim, o script resulta em uma planilha com os dados do nome do metadado, nome da instância referente a ele, o valor consultado, o identificador do objeto da Wikidata encontrado e a média da pontuação resultante do cálculo de semelhança entre os valores e os rótulos dos objetos. Com essa planilha, é possível analisar os resultados da reconciliação de dados com a Wikidata.

ANÁLISE E DISCUSSÃO DOS RESULTADOS

Os resultados apresentados a seguir constituem a descrição dos produtos da consulta dos valores da base de dados do acervo do Museu Histórico Nacional (tabela 1) na Wikidata. Serão apresentados de forma sintética os resultados da identificação de instâncias e a descrição dos objetos da Wikidata identificados em relação aos valores da base de dados do MHN. A análise dos resultados tem o objetivo de complementar o entendimento do processo de reconciliação de dados através de programação em Python, observando como os produtos dos scripts completam o ciclo de ligação de dados.

Analisando o tempo gasto para a reconciliação dos dados (tabela 2), houve uma semelhança entre os períodos dos processamentos de consulta de dados. Observando que para o metadado "Autor" houve mais tempo gasto, porém também houveram mais itens consultados, mesma condição de proporção se repete para a consulta dos valores dos demais metadados. Analisando a coluna "Média por item", que calcula a divisão entre o tempo gasto e a quantidade de itens consultados, observa-se que, independente do script, uma consulta na Wikidata leva em média de 3 a 4 segundos para ser executada, levando em conta ainda que o script adiciona um tempo de espera de 3 segundos para evitar o bloqueio de segurança da API, se não fosse essa adição de tempo, a consulta levaria um segundo ou menos para ser executada.

Esse indicador mostra a que a aplicação do script é viável com tempo de resposta considerado baixo em bases com um volume relativamente pequeno de dados, levando em conta tanto a quantidade de valores distintos que se deseja reconciliar (até 5.000) se o tempo de execução mantiver a proporção de 3 segundos por consulta.

Para afirmar com mais certeza essa viabilidade, seria interessante investir em uma análise com diferentes quantidades e tipos de conjuntos de dados, para identificar se há alguma limitação por volume e sua viabilidade em conjuntos de dados com uma grande quantidade de observações (5.000 ou mais), em que tal serviço pode ser executado com diferentes estratégias, não impactando na experiência final de navegação do usuário e permitindo um enriquecimento contínuo de acervos mesmo de grande volume.

No entanto, é importante ressaltar que essa estimativa se aplica ao contexto da reconciliação de dados em massa. Uma vez que a reconciliação seja incorporada ao sistema de indexação de objetos, ou que a instituição adote essa etapa de enriquecimento como parte da indexação de objetos digitais na web, espera-se um volume de dados mais baixo e contínuo, abrindo espaço para que o usuário valide os resultados da reconciliação por exemplo.

Tabela 2 – Tempo de execução dos scripts de reconciliação de dados

Script	Metadado	Tempo de verificação	Média por item
Identificação de instâncias	Autor	00:14:30	00:00:04
Reconciliação de dados	Autor	00:14:28	00:00:04
Identificação de instâncias	Técnica	00:02:30	00:00:03
Reconciliação de dados	Técnica	00:02:21	00:00:03
Identificação de instâncias	Material	00:02:56	00:00:03
Reconciliação de dados	Material	00:03:01	00:00:04
TOTAL		00:39:46	-

Fonte: Dados da Pesquisa (2020).

Vale ainda deixar claro que este estudo não tinha o objetivo de mensurar os níveis de viabilidade da reconciliação de dados em grande escala e, por isso, não se têm resultados suficientes para declarar qual infraestrutura tecnológica melhor se adequaria a contextos com uma quantidade de dados maior.

Observando os resultados do script de identificação de instâncias (tabela 3), um volume baixo de instâncias foi identificado para os metadados "Material" e "Técnica", sendo que este último não apresentou nenhuma instância que ocorreu mais de uma vez, já o metadado "Autor" apresentou instâncias com ocorrências maiores, as ocorrências indicam inicialmente que houve mais facilidade de identificar objetos na Wikidata com rótulos semelhantes aos valores de "Autor" do que de "Material" e "Técnica".

Tabela 3 – Resultados do script de identificação de instâncias

Metadado	Instâncias Propostas	Ocorrências
Material	ser humano	4
	elemento químico	4
	táxon	2
	material	2
	fibra	2
Autor	ser humano	205
Metadado	Instâncias Propostas	Ocorrências
Autor	artigo científico	121
	sobrenome	10
	ortsteil	7
	obra criativa	7
Técnica	visual arts technique	1
	setor econômico	1
	Página web	1
	técnica artística	1
	atividade	1

Fonte: Dados da pesquisa (2020).

Outro ponto interessante é observado nas instâncias sugeridas, como no caso do metadado "Material", com 4 ocorrências da instância "ser humano", que, inicialmente, não tem relação com os valores desse metadado, já os demais possuem algum tipo de relação. Vale destacar que somente a instância "material", que foi a selecionada, aponta relação explícita, porém só ocorreu 2 vezes. Quanto ao metadado "Técnica", somente as instâncias "visual arts technique" e "técnica artística" (selecionada) têm relação efetiva com os valores do metadado, porém só ocorreram uma vez, o que já indica uma baixa conexão com os dados da Wikidata. Isso evidencia áreas da base de conhecimento que precisam ser potencialmente melhoradas e complementadas com novos conjuntos de dados. Por último, o metadado "Autor" apresenta inicialmente duas instâncias relacionadas com os valores do metadado, "ser humano" e "sobrenome", sendo o primeiro com maior incidência, foi o escolhido para representar o metadado.

Esses resultados já indicam um nível baixo de reconhecimento dos valores dos metadados "Material" e "Técnica" na Wikidata e um nível mais alto para o metadado "Autor", o que já aponta uma maior recuperação de objetos que se relacionem com os valores de autor que os demais metadados. Isso sugere o questionamento para identificar a causa desses resultados, se a Wikidata realmente tem uma lacuna de objetos na língua portuguesa sobre esses tipos de metadados, ou se existe um arranjo melhor para identificar os objetos. Explorar bases de conhecimento dessa maneira pode gerar panoramas de áreas de conhecimento nos quais uma base pode ser melhor aplicada em relação a outras, permitindo avançar o conhecimento em seu potencial de uso.

Já observando os resultados do script de reconciliação de objetos, as indicações constatadas acima são confirmadas, como apresenta a tabela 4. O metadado "Autor" apresentou uma maior proporção de objetos identificados, diferente da proporção de identificação dos metadados "Material" e "Técnica", com um resultado consideravelmente menor.

Tabela 4 – Resultados do script de reconciliação de objetos

Metadado	Nº de Itens Distintos	Nº de Itens com Objetos Identificados	Proporção
Autor	243	157	64,61%
Material	51	2	3,92%
Técnica	43	1	2,33%

Fonte: Dados da pesquisa (2020).

Este estudo não produziu dados suficiente para identificar o motivo da baixa proporção de objetos da Wikidata identificados para os metadados "Material" e "Técnica", o que abre a premissa para a investigação dessa condição em novos estudos, cujo objetivo seja verificar, por exemplo, quais motivos levaram a este resultado.

Em busca de verificar a qualidade dos valores ligados a objetos da Wikidata, uma validação dos objetos foi realizada, identificando quais objetos realmente tinham relação com os valores consultados. Para o metadado "Técnica", o único objeto identificado foi relativo ao valor "Pintura a óleo", identificado na Wikidata como o "oil painting" de identificador "Q56676227", constituindo uma relação válida. Já no metadado "Material", os dois valores com objetos identificados na Wikidata efetivaram relações válidas: o valor "vidro" com o objeto "glass" de identificador "Q11469" e o valor "alumínio" com o objeto "aluminum" de identificador "Q663".

No metadado "Autor", dos 157 valores identificados, 143 (91,08%) deles tiveram a relação válida com os objetos da Wikidata, os demais 14 não obtiveram resultados válidos. A validade no caso deste metadado foi verificada analisando qualitativamente o objeto e identificando se o mesmo contém informações que indicam de maneira positiva a referência ao valor consultado, por exemplo, pessoas com "ocupações" relacionadas à cultura, como pintores, ilustradores ou compositores. É importante ressaltar que a taxa de acerto apresenta um resultado bastante significativo para essa instância, mostrando um potencial a ser explorado para serviços informacionais. Outro ponto interessante a ser observado é a relação entre a pontuação de semelhança dos termos do valor consultado com o rótulo do objeto da Wikidata e a validade dos objetos. O objetivo de se calcular a semelhança entre os termos foi de indicar qual objeto tem a maior possibilidade de manter uma relação válida com o valor consultado, mas no caso dos valores do metadado "Autor", a correlação entre essas duas variáveis foi muito baixa, de 0,30[18].

Uma nova pesquisa pode investigar melhor as condições de aproximação entre os valores consultados e a validade das relações com os objetos da Wikidata, se a aproximação textual é o caminho mais viável e quais são as outras possibilidades. É importante ressaltar que identificar heurísticas mais adequadas para validar esses resultados de forma automática ou semiautomática é um objetivo de pesquisa a ser explorado futuramente.

Ainda, algumas observações do processo de validação são interessantes de serem elencadas, como a importância de um preenchimento consistente dos valores. Em muitos casos, o valor do nome do autor continha somente um sobrenome ou um só nome e não o nome completo, o que dificulta a busca. Por exemplo, no caso do autor de nome "Otto" 14 objetos da Wikidata foram indicados, mas nenhum foi validado, já para os nomes completos, a ocorrência de objetos únicos e válidos foi maior.

Outra observação relevante é a capacidade do processo da Wikidata de desambiguação de valores, uma vez que a base não foi previamente normalizada, alguns nomes iguais, mas digitados de maneira diferente foram identificados como o mesmo objeto na Wikidata, o que, no final das contas, desambigua a base por si. Por exemplo, os nomes "Spanyi Ernest César Novak" e "Spanyi Ernesto César Agostinho Novak" são o mesmo autor, porém escritos de forma diferente na base de dados, ao reconhecer o mesmo objeto da Wikidata para os dois valores, este autor é desambiguado. Essa questão suscita a discussão sobre a importância da normalização dos dados e da aplicação de regras de catalogação consistentes em um processo de abertura e ligação de dados.

Um último ponto interessante é a capacidade importante que a reconciliação de dados com a Wikidata tem de reconhecer os valores consultados. Como cada objeto na Wikidata tem um rótulo e a possibilidade de rótulos alternativos, um objeto pode ser reconhecido através de suas variações terminológicas. Assim, objetos com nomes diferentes em determinadas ocasiões podem ser reconciliados e identificados como o mesmo objeto.

[18] Correlação de Pearson, calculada entre a matriz de pontuação da semelhança de termos e a matriz de validação dos objetos da Wikidata.

Por exemplo, o nome de autor "Serrano", cujo rótulo do objeto na Wikidata é "Giovanni Battista Crespi", porém pode ser reconhecido nos rótulos alternativos como "Serrano". Essa condição permite a ligação de dados entre diferentes provedores, sem perder a identidade do valor na base de origem e conservando uma única identidade ao objeto.

CONCLUSÕES

Ao refletir sobre o processo de desenvolvimento dos scripts de reconciliação de dados e seus resultados, alguns tópicos importantes demandam mais atenção:

- É nítido que um tratamento prévio dos dados e o cuidado com as regras de catalogação podem aumentar o potencial da reconciliação de dados. Como apontado nos resultados acima, algumas características indicavam a falta de normalização da base de dados, como valores vazios, valores que são iguais escritos de maneira diferentes, e, no caso dos autores, nomes incompletos. Essas condições dificultam a busca pelos termos, quanto mais completo e menos ambíguo um valor é, mais fácil será de encontrar um objeto digital correspondente em um sistema de organização do conhecimento como a Wikidata.

- Uma questão que paira ao observar a proporção de objetos efetivamente reconciliados dos metadados de "Material" e "Técnica" é o quanto a Wikidata pode ser entendida como referência para efetivar a ligação de dados com objetos digitais culturais na língua portuguesa? De acordo com os resultados, é clara a relevância da Wikidata ao buscar pelos autores, mas quanto aos outros metadados, seria necessária uma análise mais profunda das causas do baixo retorno obtido, e que confirme quais alternativas ou outras formas de abordagem podem ser elaboradas para reconciliar esses tipos de dados.

- Outro ponto de relevância indiscutível são os benefícios de se reconciliar os dados com sistemas digitais de organização do conhecimento como a Wikidata. O primeiro motivo é o auxílio à normalização dos dados com a desambiguação de termos, como cada objeto na Wikidata tem uma lista de rótulos alternativos, as diferentes variações de escrita são atendidas e, se por acaso um mesmo item aparece escrito de duas maneiras diferentes na base de dados de origem, eles podem estar contemplados no conjunto de rótulos alternativos, e serão entendidos como um único objeto digital. Outro benefício importante é o enriquecimento de dados, uma vez que os dados estão ligados, informações que não foram contempladas na base de dados de origem podem ser recuperadas da Wikidata, auxiliando a contextualizar melhor os objetos culturais, permitindo filtros por categorias antes impensáveis, como por exemplo, itens por nacionalidade do autor, ou itens cujos autores são escritores e, ainda, viabilizando a conexão com outros acervos culturais através do indicador de identificadores externos, que apontam a existência de determinado objeto em outros provedores de dados.

Entende-se, então, que o processo de reconciliação de dados, envolve a qualidade dos dados de origem, a conexão com um sistema digital de organização de conhecimento que permite consultas, formas de elencar quais resultados são mais relevantes, como a frequência das instâncias e a pontuação por semelhança entre termos e uma validação dos resultados obtidos, para, efetivamente, realizar a ligação dos dados. Cada etapa deste processo pode derivar um aprofundamento específico em busca de melhorar a forma como é utilizada e auxiliar na efetivação do processo como um todo, em busca de promover a socialização da informação que uma vez existe isolada em repositórios e portais institucionais. Promover a reconciliação dos dados, então, além de enriquecer o acervo de origem e promover a ideia dos dados abertos ligados, proporciona à sociedade seu direito de acesso à informação e, além disso, à suas raízes culturais.

Usar o meio web para promover esse tipo de fenômeno é dar ao público o que o pertence, de maneira contextualizada e interconectada, favorecendo o compartilhamento de conhecimento e desenvolvendo a interação social através da cultura digital em rede.

Dessa forma, entende-se que o processo de desenvolvimento e descrição dos scripts de reconciliação de dados atende ao objetivo proposto inicialmente, o de contribuir para o entendimento de como se dá a aplicação de uma técnica de enriquecimento semântico em bases de dados culturais, e responde ao anseio de entender como se dá o processo de reconciliação de dados culturais com a Wikidata através da programação de scripts na linguagem Python. Ao fim do desenvolvimento e experimentação dos scripts, conclui-se que eles podem ser aplicados futuramente em outros contexto de reconciliação de conjuntos de dados, em busca de fundamentar a ligação e a abertura dos dados, bem como mediar o enriquecimento da base de dados de origem, uma vez que dada a efetivação da ligação de dados, dados do sistema de organização do conhecimento (Wikidata) podem ser adicionados ao conjunto de origem, e o processo reverso também pode ocorrer, dados do conjunto de origem podem ser inseridos na Wikidata se já não existirem.

REFERÊNCIAS

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. *Scientific American*, v. 284, n. 5, p. 34-43, 2001.

BERNERS-LEE, T. *Linked data principles,* 2006. Disponível em: <http://www.w3.org/DesignIssues/LinkedData.html> Acesso em: 10 set. 2020.

ISAAC, A.; MANGUINHAS, H.; STILLER, J.; CHARLES, V. *Report on enrichment and evaluation.* The Hague, Netherlands: Europeana Task Force on Enrichment and Evaluation, 2015. Disponível em: <http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/Enrichment_Evaluation/FinalReport_EnrichmentEvaluation_102015.pdf>. Acesso em 15 de abr. de 2020.

RUBERTO, D.L.V.G.; ANTONIAZZI, R. L. Análise e Comparação de Algoritmos de Similaridade e Distância entre strings Adaptados ao Português Brasileiro. In: ANAIS DA XIII ESCOLA REGIONAL DE BANCO DE DADOS, XIII, 2017, [s.l]. *Anais…* [s.l]: SBC, 2017.

SANDERSON, R. *"The Linked Data Snowball and Why We Need Reconciliation"*, 2016. Disponível em: <https://www.slideshare.net/azaroth42/linked-data-snowball-or-why-we-need-reconciliation.> Acesso em: 28 mar. 2020.

SANTARÉM SEGUNDO, J. E. Web Semântica: Introdução a recuperação de dados usando SPARQL. In: *Encontro Nacional de Pesquisas em Ciência da Informação (ENANCIB),* v. 14, p. 3242-3261, 2014.

SOUZA, R. R.; ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. *Ciência da Informação*, v. 33, n. 1, p. 132-141, 2004.

VRANDEČIĆ, D.; KRÖTZSCH, M. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, v. 57, n. 10, p. 78-85, 2014.

ZENG, M. L. Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article. *El profesional de la información*, v. 28, n. 1, 2019.

AGRADECIMENTOS

Agradecimentos à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por fomentar a produção deste artigo através do programa de bolsa de pesquisa para doutorandos, e ao grupo de pesquisa Laboratório de Inteligência de Redes (UnB) por fomentar e conceder infraestrutura para a aplicação desta pesquisa