



Web semântica em repositórios: ontologia para representação de bibliotecas digitais

Semantic web in repositories: ontology for representation of digital libraries

Marcel Ferrante Silva 

Doutor em Ciência da Informação
Universidade Federal de Goiás
marcelf@gmail.com

Dalton Lopes Martins 

Doutor em Ciência da Informação
Universidade de Brasília
dmartins@gmail.com

Joyce Siqueira 

Doutoranda em Ciência da Informação
Universidade de Brasília
joycitta@gmail.com

Resumo

A quantidade de objetos digitais armazenados em repositórios digitais cresce substancialmente, tornando iminente a necessidade de softwares interoperáveis e semânticos, que permitam às máquinas realizar, de forma automatizada, serviços como, processar, recuperar e reutilizar conteúdo. Para tal, é essencial o uso de ontologias que expressem as relações semânticas entre os objetos digitais. Dessa forma, este estudo propôs um modelo conceitual de ontologia voltada à representação de conteúdo semântico em repositórios digitais, utilizando como base duas consolidadas metodologias de desenvolvimento de ontologias, o Método 101, desenvolvido por Noy e McGuinness (2001) e o método de Uschold e King (1995), assim como baseou-se nos preceitos da ontologia SIOC e do DSpace. Ao final, apresentou-se um modelo conceitual da ontologia para representação do conteúdo de um repositório digital em UML. O resultado visa aprimorar os processos, principalmente, para recuperação de dados na web, possibilitando ao usuário melhores resultados em suas buscas.

Palavras-chave

Modelo conceitual. Ontologia. Repositório digital. Web semântica.

Abstract

The number of digital objects stored in digital repositories grows substantially, making imminent the need for interoperable and semantic software that allows machines to automate services such as processing, retrieving and reusing content. For this, it is essential to use ontologies that express the semantic relations between digital objects. Thus, this study proposed a conceptual ontology model for the representation of semantic content in digital repositories, based on two consolidated methodologies of ontology development, Method 101, developed by Noy and McGuinness (2001) and the Uschold method King (1995), as well as, was based on the precepts of the SIOC ontology. At the end, a conceptual model of the ontology was presented to represent the contents of a digital repository in UML. The result is to improve the processes, mainly for data recovery on the web, allowing the user better results in their searches.

Keywords

Conceptual model. Ontology. Digital repository. Semantic web.



1 INTRODUÇÃO

A Web Semântica refere-se a um conjunto de tecnologias que visam unir semanticamente objetos via web, ou seja, que permitam agregar informações contextualizadas sobre o objeto, produzindo arquivos inteligíveis a agentes computacionais, possibilitando buscas automatizadas mais efetivas e resultados mais precisos.

Para que os repositórios digitais possam contribuir e usufruir da Web Semântica, seus dados, até então guardados unicamente em banco de dados relacionais, devem se tornar dados abertos em formato *Resource Description Framework* (RDF). Além disso, é necessário utilizar uma ou mais ontologias para definir o significado dos termos usados nas declarações RDF. Em essência, o RDF e as ontologias fornecem contexto e significado a uma declaração.

Nesse sentido, esta pesquisa tem o objetivo de desenvolver uma ontologia própria a ser usada para interoperabilidade semântica entre repositórios digitais. Para tal, inicia-se com alguns questionamentos: Qual modelo conceitual deve ser seguido? Qual linguagem deve ser utilizada? Quais entidades devem ser representadas? Quais devem ser seus atributos e relacionamentos? Além disso, ainda é necessário considerar o perfil do usuário, criado pelo advento da Web 2.0 e dos paradigmas propostos pelo uso das mídias sociais.

A partir do estudo destas questões, esta pesquisa desenvolveu uma ontologia para representação de um repositório digital e seus elementos. A ontologia foi elaborada utilizando o formato RDF e *Ontology Web Language* (OWL). Sendo assim, os objetivos deste estudo são: a) primeiramente, definir formalmente um repositório e seus elementos através de uma ontologia; b) servir de referência para um sistema de gestão repositório digital exportar semanticamente seu conteúdo, possibilitando o reuso da informação por outros sistemas de informação; e c) criar condições para construção de uma rede de repositórios com busca centralizada que compartilhe de um mesmo vocabulário para intercâmbio de informações sobre o conteúdo de cada repositório, independente da fabricante do software; e c) contribuir para a discussão de um modelo de bibliotecas e repositórios digitais no campo da ciência da informação, levando em consideração a web semântica.

Nas próximas seções são apresentados conceitos importantes para compreensão do tema, seguido da metodologia utilizada para criar a ontologia e dos resultados obtidos.

2 FUNDAMENTOS TEÓRICOS

As tecnologias da web evoluem em prol de propiciar aos usuários experiências cada vez mais úteis e imersivas. A Web 1.0 ou Web de Documentos, para Laufer (2015), era composta basicamente por documentos estáticos, com o propósito de tornar público conteúdos, organizados em hipertextos. A partir das novas necessidades, surgiu a Web Programável, na qual aplicações utilizam os navegadores para a execução de programas. Na Web Programável, a interoperabilidade passa a ser um ponto importante, visto a necessidade de uma aplicação “conversar” com outra.

Pode-se compreender interoperabilidade como a propriedade de sistemas diferentes [...], através de padrões tecnológicos, acordos ou propostas, de serem capazes de operar em conjunto, visando a execução de uma tarefa (MARCONDES, 2016, p. 68).

De acordo com Gil (2014), a Web 2.0 ou Web Social nasceu a partir do desenvolvimento de ferramentas que possibilitam a interação entre usuários, tais como blogs e redes

sociais, ou seja, que permitem relações socioafetivas e a correspondente humanização das relações via dispositivos digitais. Estas interações, com discussão e reflexão crítica, passaram-se a criar condições para a 3ª geração, a Web 3.0 ou Web Semântica.

Para Catarino, Cervantes e Almeida (2015), a Web Semântica propõe a representação de dados na web, por meio de padrões propostos pela *World Wide Web Consortium (W3C)*, com o intuito de agregar ao dado significado que considere seu contexto. Por consequência, visa uma nova proposta para a Web, na qual as informações devem se tornar legíveis às máquinas, por meio da aplicação de diferentes tecnologias, linguagens, padrões e recomendações.

O projeto concebe a ideia de uma web de dados vinculados, *Linked Data*, disponibilizados em formato padrão, acessível e gerenciável.

O conceito de Web Semântica e de ligação entre dados deu origem ao *Linked Data*, um conjunto de dados em formato aberto, que se apresenta como um grande ambiente de informação através de dados ligados semanticamente entre si em vários lugares ao redor do mundo, caracterizando-se por utilizar a internet (a nuvem) como seu hospedeiro. É possível afirmar que o *Linked Data* é uma materialização dos conceitos da Web Semântica (SANTARÉM SEGUNDO, 2014, p. 3243).

Wood (2014) justifica que no cenário atual da Web é possível acessar uma infinidade de dados, com diferentes formatos, PDF, TIFF, CSV, planilhas eletrônicas, tabelas incorporadas a documentos de texto e muitas outras formas de texto simples. Todos eles bem diferentes em essência, mas com um fator em comum: são formatados para o consumo humano, ou seja, para extrair qualquer informação é preciso análise manual. Não obstante, ainda há uma dificuldade latente de processos automatizados que acessem, pesquisem ou reutilizem esses dados.

Para Wood (2014), a *Linked Data* tem por objetivo transformar a Web em um banco de dados global, o que tornaria possível a usuários especialistas consultar dados de várias fontes e combiná-los em tempo real. Refere-se a um conjunto composto por melhores práticas para publicar e conectar dados estruturados na Web, utilizando padrões internacionais da W3C.

Retornando ao cenário supracitado, no qual há uma infinidade de dados, em diversos formatos, Tim Berners-Lee sugeriu, para orientar a publicação de dados abertos, um esquema de classificação chamado “5 Estrelas para Dados Abertos” (*5 Stars Open Data*), onde é traçado um caminho evolutivo na direção desse universo de dados conectados. Este caminho está apresentado na Figura 1.

Figura 1 - 5 Estrelas para Dados Abertos



Fonte: Adaptada de *5 Stars Open Data* (2018).

De acordo com Santarém Segundo (2014), para concretizar o conceito de dados abertos, tecnologias como *Resource Description Framework (RDF)*, *eXtensible Markup Language (XML)* e *Web Ontology Language (OWL)* tornam possível o processo de construção e armazenamento da informação, constituindo assim ambientes que possam ter conjunto de dados ligados semanticamente.

Para esta pesquisa, dá-se destaque ao RDF, que é citada pela W3C (2018) e *W3C Recommendation* (2004), como uma estrutura escrita em XML, que descreve recursos na web, de forma a ser lido e entendido exclusivamente por computadores, ou ainda, um padrão para representar informações sobre dados na Web, destinado a representar metadados, que forneçam uma estrutura comum para expressar informações, sem perda de conteúdo, de forma interoperável.

Baseia-se na concepção de que os dados possuem propriedades, que possuem valores e que os recursos podem ser descritos a partir de declarações, segundo a *W3Schools* (2018). O Quadro 1 mostra um exemplo de código RDF.

Quadro 1 - Código RDF referente ao recurso “CD Empire Burlesque”

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:cd="http://www.recshop.fake/cd#">
<rdf:Description rdf:about="http://www.recshop.fake/cd/Empire Burlesque">
<cd:artist>Bob Dylan</cd:artist>
<cd:company>Columbia</cd:company>
<cd:price>10.90</cd:price>
<cd:year>1985</cd:year>
</rdf:Description>
</rdf:RDF>
```

Fonte: Adaptado de *W3Schools* (2018).

Para Laufer (2015), a Web Semântica se relacionada à forma como os metadados são agregados às informações e é fundamental que o significado pretendido pelo publicador dos dados seja o mesmo que o entendido pelo consumidor dos dados, por isso é imprescindível o uso de ontologias com uma semântica bem definida.

Segundo Carlan (2010), uma ontologia define quais termos devem ser usados para descrever e representar uma área do conhecimento, podendo ser usada por pessoas, bancos de dados, em técnicas e aplicações de raciocínio indutivo e inferências. Borst (1997) conceitua ontologia como uma especificação formal (legível para os computadores), e explícita (com conceitos, propriedades, relações, funções, restrições e axiomas explicitamente definidos) de uma conceitualização (representa um modelo abstrato de algum fenômeno do mundo real) compartilhada (um conhecimento consensual entre os pares).

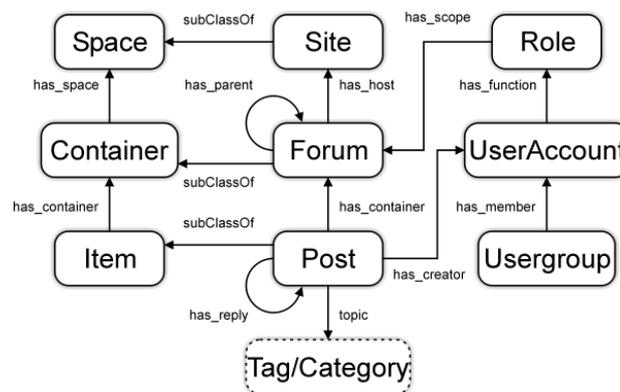
Há diferentes ontologias de referência como, por exemplo, *Friend of a Friend (FOAF)*, *Simple Knowledge Organization System (SKOS)* e *Semantically-Interlinked Online Communities (SIOC)*, dentre outras. Cada ontologia é descrita por um documento apontado por uma URI ou *namespace* e as classes e as propriedades são construídas a partir da concatenação da URI com o nome da classe ou propriedade desejada. Na ontologia FOAF, disponível em FOAF (2014), por exemplo, o *namespace* é “<http://xmlns.com/foaf/0.1/>” e uma das propriedades é a *name* “<http://xmlns.com/foaf/0.1/:name>”.

Caso as ontologias disponíveis não atendam, pode-se criar novas, sempre reutilizando elementos das ontologias de referência, conforme Laufer (2015). Considerando a criação de novas ontologias, esta pesquisa enfatiza a OWL, que de acordo com a *W3C Recommendation*

(2004), se trata de uma linguagem para definir e instanciar ontologias da Web, com descrições de classes, propriedades e suas instâncias, compreendidas por máquinas. Esse esforço visa tornar os recursos da Web mais acessíveis para processos automatizados, adicionando informações sobre os recursos que descrevem ou fornecem conteúdo da Web.

Como supracitado, é importante reutilizar elementos das ontologias de referência. Para esta pesquisa o SIOC teve papel primordial, por ter como foco as redes sociais. Segundo a W3C (2016) e Breslin *et al.* (2009), o projeto SIOC, traduzido para “Comunidades Online Semanticamente Interligadas”, visa modelar a Web Social, usando tecnologias da Web Semântica, a partir da descrição das informações contidas nos *sites* de comunidades online: blogs, fóruns, wikis etc., visto que o uso extensivo das redes sociais despertou novas vertentes em relação à publicação e ao compartilhamento de dados, o que suscitou questões relacionadas à interoperabilidade e ao intercâmbio de contribuições de entre as aplicações. A Figura 2 apresenta a visão geral do SIOC.

Figura 2 - Visão Geral do SIOC



Fonte: Berrueta *et al.* (2018).

Breslin *et al.* (2009) afirmam ainda que o projeto objetivou, ao desenvolver a ontologia SIOC, propor uma solução simples, no entanto, robusta, de modo que pudesse ser facilmente implantada em aplicações existentes. O SIOC faz uso do padrão *Dublin Core* para definição de vários atributos de conteúdo e do FOAF - *Friend Of A Friend*, para modelar identidade pessoal e atributos relacionados. Desse modo, embora haja diferentes perfis de usuários em diferentes *sites*, todos se relacionam com a mesma pessoa física, o que permite que esta possa ser modelada em um único gráfico global de dados RDF, permitindo um modelo de interoperabilidade e portabilidade de dados sociais entre serviços.

Com a Web devidamente estruturada, há uma infinidade de dados disponíveis, surgindo a necessidade da utilização de formatos e estruturas adequadas de pesquisa. Surgindo assim, o protocolo *Simple Protocol and RDF Query Language (SPARQL)*, que, de acordo com a *W3C Recommendation (2008)*, se trata de uma linguagem de consulta, que utiliza um conjunto de especificações que fornecem linguagens e protocolos para consultar e manipular o conteúdo publicado em RDF na Web.

Segundo *SPARQL by Example (2018)*, com o SPARQL é possível extrair valores de dados estruturados (ex. bancos de dados relacionais) e semiestruturados (ex. documentos xml); explorar dados consultando relacionamentos desconhecidos; realizar junções complexas de bancos de dados diferentes em uma única consulta simples e transformar dados de RDF de um vocabulário para outro. Semelhante ao SQL, o SPARQL possui uma estrutura *Select-From-Where*.

A Web Semântica é considerada uma extensão da Web tradicional, trazendo mudanças nas buscas na Internet, que não mais contarão apenas com as muito utilizadas “palavras-chaves”, mas com palavras enriquecidas com seu significado ou significados.

Para que os repositórios digitais possam contribuir e usufruir da Web Semântica, seus dados, guardados em bancos relacionais, devem se tornar dados abertos ligados, proporcionando não apenas interoperabilidade semântica entre os repositórios, mas também com toda rede.

A maioria dos repositórios digitais relacionados com o movimento de Acesso Livre (AL) implementam o *Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)*.

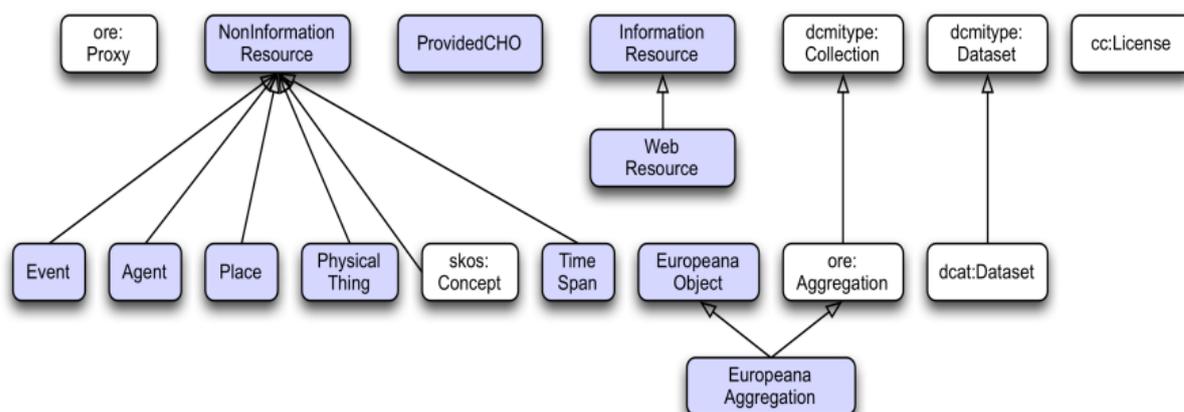
O Protocolo da *Open Archives Initiative for Metadata Harvesting (OAI-PMH)* é um mecanismo de baixa barreira para a interoperabilidade do repositório. *Data Providers* são repositórios que expõem metadados estruturados via OAI-PMH. Os provedores de serviços, em seguida, fazem solicitações de serviço OAI-PMH para coletar esses metadados (OPEN ARCHIVES INITIATIVE, 2019).

No entanto, a necessidade da especificação semântica tornou-se evidente, incitando pesquisas como a deste artigo, que procuram meios de prover a interoperabilidade semântica entre os repositórios e em segunda instância, com toda Web, neste caso, especificamente por meio do desenvolvimento de um modelo conceitual de ontologia escrita em RDF/OWL.

Um exemplo de modelo para interoperabilidade é o caso da Europeana. De acordo com Patrício (2012), uma alta quantidade de objetos pertencentes ao patrimônio cultural da Europa é digitalizada por grandes instituições, como bibliotecas e museus, e cada uma utilizava seu próprio padrão de metadados. Freire, Charles e Chambers (2012, p. 1 *apud* CAR-RASCO; VIDOTTI; MONTEIRO, 2016) afirmam que, para solucionar esse problema, a primeira iniciativa da Europeana, biblioteca virtual desenvolvida pelos países da União Europeia, foi o desenvolver o *Europeana Semantic Elements (ESE)*, que possibilita a interoperabilidade para trocas de metadados, baseado em Dublin Core.

Para Europeana Pro (2013) e Patrício (2012), o EDM é a evolução do ESE e visa a interoperabilidade com o uso da Web Semântica, por meio da especificação formal das classes e propriedades que podem ser usadas na Europeana, garantindo uma melhor preservação da riqueza dos dados, sem prejudicar a interoperabilidade. A Figura 3 apresenta a hierarquia de classes do EDM. Destaca-se que as classes introduzidas pelo EDM são mostradas em retângulos azuis e as demais são classes reutilizadas de outros esquemas.

Figura 3 - Hierarquia de Classes do EDM

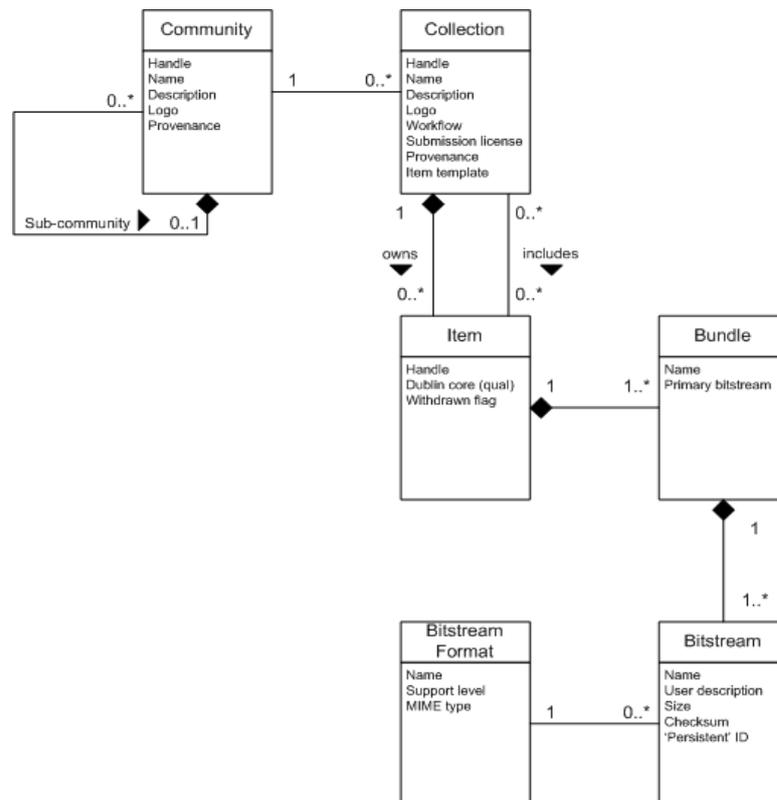


Fonte: *Europeana Think Culture* (2017, p. 7).

Europeana Pro (2013) e Patrício (2012) destacam também que a EDM adota a Web Semântica, como aberto e interdomínio que pode acomodar o alcance e a riqueza de padrões comunitários específicos, como LIDO para museus, EAD para arquivos ou METS para bibliotecas digitais. Dessa forma, possibilita que cada fornecedor da Europeana estruture os seus dados sem que a respectiva riqueza original se perca.

Outra referência para o estudo foi o sistema de gestão de repositório digital DSpace, software livre amplamente utilizado por instituições de ensino no Brasil e no mundo. Seu modelo de dados, definido utilizando a linguagem de modelagem *Unified Modeling Language (UML)*, é apresentado na Figura 4.

Figura 4 - Diagrama de objetos do DSpace



Fonte: Gorton (2007, p. 18).

O diagrama apresenta no topo as comunidades que têm alguns atributos como nome, logo, descrição, etc. As comunidades (*community*) podem ter sub-comunidades e nelas são associados os usuários. As comunidades são constituídas por coleções (*collection*), mas as coleções podem estar em várias comunidades. Uma coleção possui itens. Um item possui atributos definidos por um esquema de metadados (no diagrama é mencionado o Dublin core) e é formado pelo conjunto de objetos digitais (*bitstream*) que possuem diferentes formatos (*bitstream format*).

3 METODOLOGIA

Para Fernandez, Gómez Perez e Juristo (1997), apesar de diversas ontologias terem sido desenvolvidas por diferentes comunidades, sob diferentes abordagens, métodos e técnicas, ainda não existe um consenso em relação a melhor metodologia para seu processo de

construção. O que acarreta na ausência de atividades padronizadas, cujas realizações são conduzidas de forma artesanal, não científica. Verifica-se ainda a falta de explicação sistemática de como e onde serão usadas as abordagens teóricas dentro de seu processo de elaboração.

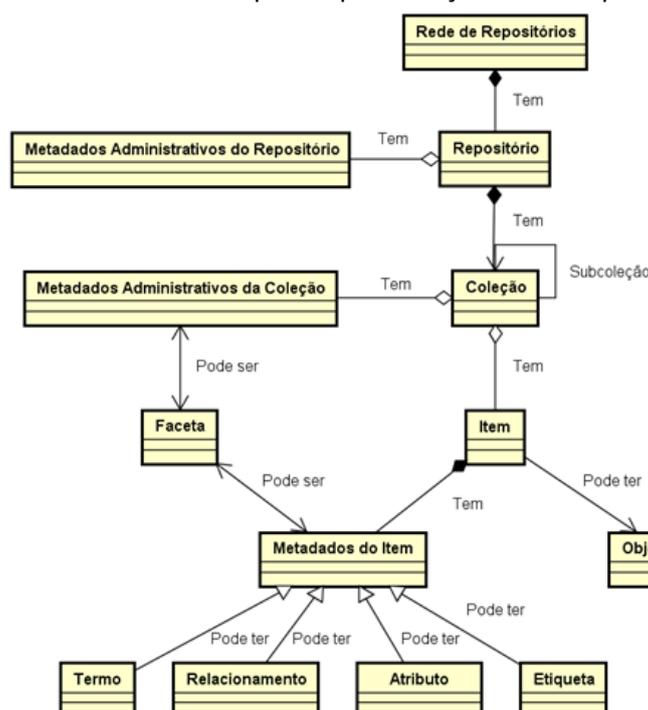
Desse modo, foram utilizadas de forma combinada as seguintes metodologias para construção da ontologia: a primeira, proposta por Natalya F. Noy e Deborah L. McGuinness (2001), intitulada Método 101, foi extensivamente utilizada. O método descreve uma abordagem iterativa para o desenvolvimento da ontologia, cujo processo se divide nas seguintes etapas: 1. Determinar o domínio e o escopo da ontologia; 2. Considerar ontologias existentes; 3. Enumerar termos importantes na ontologia; 4. Definir as classes e sua hierarquia; 5. Definir as propriedades das classes; 6. Definir as restrições e 7. Criar instâncias; a segunda metodologia, proposta por Mike Uschold e Martin King (1995), devem ser considerados os seguintes itens, para criação de uma ontologia: 1. Identificar o propósito da ontologia, ou seja, a necessidade de construção, o grau de formalismo e as classes de usuários; 2. Conceber a conceitualização da ontologia, sua codificação por meio de uma linguagem de representação de ontologias e integrar com ontologias existentes; 3. Avaliar a ontologia por meio dos requisitos especificados; e, por último 4. Documentar acerca das pretensões da ontologia e das primitivas usadas para expressar as definições na ontologia.

Na seção seguinte, é apresentado o produto dessas metodologias na elaboração da representação do domínio das bibliotecas digitais.

4 ONTOLOGIA PARA REPRESENTAÇÃO DE BIBLIOTECAS DIGITAIS EM OWL

Primeiramente, para um melhor conhecimento do domínio, foi elaborado o diagrama de classes da linguagem de modelagem UML para uma representação visual do repositório e seus elementos, como apresenta a Figura 5.

Figura 5 - Diagrama de classes UML para representação de um repositório e seus elementos



Fonte: Elaborada pelos autores (2018).

Depois, utilizou-se as metodologias apresentadas na seção anterior para elaboração da ontologia. Os estudos apresentados nos Fundamentos Teóricos serviram como inspiração, principalmente o modelo de dados do sistema DSpace, que revelou muitas entidades do domínio e o SIOC, ontologia voltada para comunidades online e fóruns representarem o seu conteúdo, tem objetivo similar ao desejado neste estudo, com a diferença que este é destinado aos repositórios.

O modelo considera o repositório e seus elementos como ponto de partida para a modelagem. Para cada repositório está prevista a inclusão de metadados administrativos, além de coleções e, quando necessário, subcoleções, para as quais também está prevista a inclusão de metadados administrativos, com o diferencial, que estes metadados podem ser selecionados para configuração de facetas.

Populando as coleções e subcoleções, os itens podem ter o objeto digital original incluído ou apenas os metadados do item, que também podem ser selecionados para serem usados como facetas, principalmente na busca, atuando como filtros. Os metadados dos itens podem ser compostos por Termos, Relacionamento, Atributo e Etiqueta.

Para melhor compreensão, apresenta-se os conceitos das classes contidas no modelo, com exceção, Rede de Repositório e Repositório, descritos anteriormente:

- a) *Metadados Administrativos*: fornecem informações que apoiam a gestão dos recursos registrando as especificidades e dependências técnicas do recurso. Exemplos de metadados administrativos são as permissões dos usuários, a logo ou o “dono” da coleção.
- b) *Coleção*: conjunto de itens agrupados por uma determinada temática ou finalidade e organizados segundo algum esquema de metadados e sistema de organização. As coleções podem ter subcoleções.
- c) *Faceta*: forma alternativa de agrupar conhecimento por temas pré-estabelecidos, de forma a facilitar a interação do usuário com o conteúdo disponível.
- d) *Item*: formado pelo objeto digital ou apenas pelos metadados de descrição, que também podem apontar para o objeto de conteúdo original em outro repositório.
- e) *Objeto original*: o próprio objeto digital, que pode ou não ser armazenado.
- f) *Metadados do item*: tem a finalidade de representar o item na coleção ou subcoleções, para organizá-lo e para recuperá-lo. São eles:
 - *Termo*: faz parte de um vocabulário controlado utilizado na descrição de um item. Pode estar ordenado em uma lista de termos alfabeticamente ou pode ser estruturado hierarquicamente constituindo uma taxonomia.
 - *Relacionamento*: responsável por definir vínculos entre instâncias, relacionando-os. Por exemplo, duas coleções diferentes, livros e autores, os livros podem ser relacionados a seus respectivos autores ou vice-versa.
 - *Etiqueta*: trata-se de um termo digitado livremente pelo usuário para a descrição de um item sem um controle ou normalização.
 - *Atributo*: recebe texto que pode ser digitado livremente pelo usuário.

Termo, etiqueta e atributo, ainda que possuam semelhanças, diferem no fato de o termo possuir hierarquia e integrar um vocabulário controlado; sendo assim, o usuário na indexação e na busca escolhe entre as opções fornecidas; a etiqueta tem digitação livre pelos usuários, sendo considerada um vocabulário descontrolado, permite a colaboração e o reuso das etiquetas na indexação e na busca. E, por fim, o atributo, que apenas define a relação entre um item e um dado ou sequência de caracteres.

O código RDF/OWL da Ontologia para Representação de Bibliotecas Digitais ou *Digi-*

tal Libraries Representation Ontology (DLRO) é apresentada no Quadro 2.

Quadro 2 - Ontologia para Representação de Bibliotecas Digitais (continua)

```

<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#">

  <owl:Class rdf:about="#Repository">
    <rdfs:label xml:lang="en">Repository</rdfs:label>
    <rdfs:label xml:lang="pt">Repositório</rdfs:label>
  </owl:Class>

  <owl:Class rdf:about="#Collection">
    <rdfs:label xml:lang="en">Collection</rdfs:label>
    <rdfs:label xml:lang="pt">Coleção</rdfs:label>
  </owl:Class>

  <owl:Class rdf:about="#Item">
    <rdfs:label xml:lang="en">Item</rdfs:label>
    <rdfs:label xml:lang="pt">Item</rdfs:label>
  </owl:Class>

  <owl:Class rdf:about="#RepositoryMetadata">
    <rdfs:label xml:lang="en">RepositoryMetadata</rdfs:label>
    <rdfs:label xml:lang="pt">MetadadoDoRepositório</rdfs:label>
  </owl:Class>

  <owl:Class rdf:about="#CollectionMetadata">
    <rdfs:label xml:lang="en">CollectionMetadata</rdfs:label>
    <rdfs:label xml:lang="pt">MetadadoDaColeção</rdfs:label>
  </owl:Class>

  <owl:Class rdf:about="#Facet">
    <rdfs:label xml:lang="en">Facet</rdfs:label>
    <rdfs:label xml:lang="pt">Faceta</rdfs:label>
    <rdfs:subClassOf rdf:about="#CollectionMetadata">
  </owl:Class>

  <owl:Class rdf:about="#ItemMetadata">
    <rdfs:label xml:lang="en">ItemMetadata</rdfs:label>
    <rdfs:label xml:lang="pt">MetadadoDoItem</rdfs:label>
  </owl:Class>

  <owl:Class rdf:about="#Term">
    <rdfs:label xml:lang="en">Term</rdfs:label>
    <rdfs:label xml:lang="pt">Termo</rdfs:label>
    <rdfs:subClassOf rdf:resource="#ItemMetadata"/>
  </owl:Class>

  <owl:Class rdf:about="#Relationship">
    <rdfs:label xml:lang="en">Relationship</rdfs:label>
    <rdfs:label xml:lang="pt">Relacionamento</rdfs:label>
    <rdfs:subClassOf rdf:resource="#ItemMetadata"/>
  </owl:Class>

  <owl:Class rdf:about="#Attribute">
    <rdfs:label xml:lang="en">Attribute</rdfs:label>
    <rdfs:label xml:lang="pt">Atributo</rdfs:label>
    <rdfs:subClassOf rdf:resource="#ItemMetadata"/>
  </owl:Class>

  <owl:Class rdf:about="#Tag">
    <rdfs:label xml:lang="en">Tag</rdfs:label>
    <rdfs:label xml:lang="pt">Etiqueta</rdfs:label>
    <rdfs:subClassOf rdf:resource="#ItemMetadata"/>
  </owl:Class>

```

Fonte: Elaborado pelos autores (2018).

Quadro 2 - Ontologia para Representação de Bibliotecas Digitais (continuação/continua)

```

<owl:Class rdf:about="#DigitalObject">
  <rdfs:label xml:lang="en">DigitalObject</rdfs:label>
  <rdfs:label xml:lang="pt">ObjetoDigital</rdfs:label>
</owl:Class>

<!-- Object Properties -->

<owl:ObjectProperty rdf:about="#belongCollection">
  <rdfs:label xml:lang="en">belongCollection</rdfs:label>
  <rdfs:label xml:lang="pt">pertenceColeção</rdfs:label>
  <rdfs:range rdf:resource="#Collection"/>
  <rdfs:domain rdf:resource="#Item"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#belongRepository">
  <rdfs:label xml:lang="en">belongRepository</rdfs:label>
  <rdfs:label xml:lang="pt">pertenceRepositório</rdfs:label>
  <rdfs:domain rdf:resource="#Collection"/>
  <rdfs:range rdf:resource="#Repository"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#hasCollection">
  <rdfs:label xml:lang="en">hasCollection</rdfs:label>
  <rdfs:label xml:lang="pt">temColeção</rdfs:label>
  <rdfs:range rdf:resource="#Collection"/>
  <rdfs:domain rdf:resource="#Repository"/>
  <owl:inverseOf rdf:resource="#belongRepository"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#hasCollectionParent">
  <rdfs:range rdf:resource="#Collection"/>
  <rdfs:domain rdf:resource="#Collection"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#hasItem">
  <rdfs:label xml:lang="en">hasItem</rdfs:label>
  <rdfs:label xml:lang="pt">temItem</rdfs:label>
  <rdfs:domain rdf:resource="#Collection"/>
  <rdfs:range rdf:resource="#Item"/>
  <owl:inverseOf rdf:resource="#belongCollection"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#hasObject">
  <rdfs:label xml:lang="en">hasObject</rdfs:label>
  <rdfs:label xml:lang="pt">temObjeto</rdfs:label>
  <rdfs:domain rdf:resource="#Item"/>
  <rdfs:range rdf:resource="#Object"/>
  <owl:inverseOf rdf:resource="#belongCollection"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#hasMetadata">
  <rdfs:label xml:lang="en">hasMetadata</rdfs:label>
  <rdfs:label xml:lang="pt">temMetadado</rdfs:label>
  <rdfs:domain rdf:resource="#Item"/>
  <rdfs:range rdf:resource="#ItemMetadata"/>
  <owl:inverseOf rdf:resource="#belongCollection"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#hasCollectionMetadata">
  <rdfs:label xml:lang="en">hasCollectionMetadata</rdfs:label>
  <rdfs:label xml:lang="pt">temMetadadoDaColeção</rdfs:label>
  <rdfs:domain rdf:resource="#Collection"/>
  <rdfs:range rdf:resource="#CollectionMetadata"/>
  <owl:inverseOf rdf:resource="#belongCollection"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#hasRepositoryMetadata">
  <rdfs:label xml:lang="en">hasRepositoryMetadata</rdfs:label>

```

Fonte: Elaborado pelos autores (2018).

Quadro 2 - Ontologia para Representação de Bibliotecas Digitais

(continuação)

```

<rdfs:label xml:lang="pt">temMetadadoDoRepositório</rdfs:label>
<rdfs:domain rdf:resource="#Repository"/>
<rdfs:range rdf:resource="#RepositoryMetadata"/>
<owl:inverseOf rdf:resource="#belongCollection"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="#usedAsFacet">
  <rdfs:label xml:lang="en">usedAsFacet</rdfs:label>
  <rdfs:label xml:lang="pt">usadoComoFaceta</rdfs:label>
  <rdfs:domain rdf:resource="#Metadata"/>
  <rdfs:range rdf:resource="#Facet"/>
  <owl:inverseOf rdf:resource="#belongCollection"/>
</owl:ObjectProperty>

</rdf:RDF>

```

Fonte: Elaborado pelos autores (2018).

O modelo conceitual foi implantado no software Tainacan, que, segundo Martins *et al.* (2017), é um sistema *web*, de código aberto, voltado à gestão de repositórios digitais, desenvolvido com base no software livre WordPress. Este repositório tem por objetivo incorporar funcionalidades que facilitem a interoperabilidade com mídias sociais e permita ampliar o grau de participação social de usuários do repositório em atividades de sua gestão e manutenção, criando um repositório de fácil utilização, configuração e implementação.

5 CONCLUSÃO

Esta pesquisa objetivou criar uma ontologia, voltada para sistemas gestores de repositórios digitais, visando a proporcionar a interoperabilidade, a partir de recursos da Web Semântica.

A ontologia criada destaca-se por representar não apenas os metadados do item, mas também os metadados do repositório e da coleção. Permite também explicitar relacionamentos entre itens de uma mesma coleção ou entre coleções diferentes, expandindo o uso das coleções em um repositório, que agora pode representar semanticamente os documentos ou livros, mas também os autores, termos, publicadores e qualquer outra entidade que houver necessidade. Nesse sentido, a ontologia é uma proposta para a modelagem de um repositório semântico que atua como uma espécie de banco de dados, sendo que a coleção atua como uma classe ou entidade e os itens como indivíduos ou registros.

A ontologia também prevê um rico esquema para representação dos itens, que além dos relacionamentos, tem os atributos que agregam dados aos itens como a data de publicação, os termos que constituem taxonomias e também as etiquetas, que é uma forma moderna de classificação advinda da Web 2.0.

Com a apresentação dos resultados da pesquisa, espera-se contribuir com o avanço da área, principalmente, inspirando novos pesquisadores a utilizar e refinar o modelo. Este também pode servir para embasar um padrão oficial para repositórios a ser publicado pela W3C, assim como é feito com SIOC, que é voltado para comunidades online. Em futuros trabalhos, pretende-se evoluir a ontologia levando em consideração outros modelos e ontologias como o vocabulário *schema.org* utilizado pela biblioteca mundial WorldCat, uso da SKOS principalmente na definição dos termos e também o padrão *Object Reuse and Exchange (ORE)* publicado pela iniciativa *openarchives.org*.

REFERÊNCIAS

- 5 STARS OPEN DATA. **5 Stars Open Data**. 2018. Disponível em: <https://5stardata.info/en/>. Acesso em: 15 abr. 2019.
- BERRUETA, D. *et al.* **SIOC Core Ontology Specification**. 2018. Disponível em: <http://rdfs.org/sioc/spec/>. Acesso em: 01 jul. 2018.
- BORST, W. N. **Construction of Engineering Ontologies for Knowledge Sharing and Reuse**. PhD thesis, University of Twente, P.O. Box 217 - 7500 AE Enschede - The Netherlands, 1997.
- BRESLIN, J. G. *et al.* **SIOC: Content Exchange and Semantic Interoperability Between Social Networks**. W3C Workshop on the Future of Social Networking, 2009, Barcelona. Disponível em: <https://www.w3.org/2008/09/msnws/papers/sioc.html>. Acesso em: 22 fev. 2018.
- CARLAN, E. **Sistemas de Organização do Conhecimento: uma reflexão no contexto da Ciência da Informação**. 2010. Tese (Doutorado em Ciência da Informação), Programa de Pós-Graduação em Ciência da Informação, Universidade de Brasília, Brasília, DF, 2010.
- CARRASCO, L. B.; VIDOTTI, S. A. B. G.; MONTEIRO, S. D. Um olhar sobre a coleta de dados da Europeana. *In: ENCONTRO INTERNACIONAL DADOS, TECNOLOGIA E INFORMAÇÃO*, 3., 2016, Marília. **Anais...** Marília: Unesp, 2016.
- CATARINO, M. E.; CERVANTES, B. M. N.; ALMEIDA, I. A. A representação temática no contexto da web semântica. **Informação & Sociedade: Estudos**, João Pessoa, v. 25, n. 3, p. 105-116, set./dez. 2015. Disponível em: <http://www.periodicos.ufpb.br/ojs/index.php/ies/article/view/16242/14531>. Acesso em: 15 abr. 2019.
- EUROPEANA PRO. **Europeana Data Model Primer**. 2013. Disponível em: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf. Acesso em: 15 abr. 2019.
- EUROPEANA THINK CULTURE. **Definition of the Europeana Data Model v. 5.2.8**. 2017. Disponível em: https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf. Acesso em: 15 abr. 2019.
- FERNANDEZ, M.; GOMEZ-PEREZ, A.; JURISTO, H. Methontology: from ontological art towards ontological engineering. **AAAI Technical Report**, SS-97-06, 1997. Disponível em: http://oa.upm.es/5484/1/METHONTOLOGY_.pdf. Acesso em: 20 ago 2018.
- FOAF. **FOAF Vocabulary Specification 0.99**. 2014. Disponível em: <http://xmlns.com/foaf/spec/>. Acesso em: 15 abr. 2019.
- GIL, H. A passagem da Web 1.0 para a Web 2.0 e... Web 3.0: potenciais consequências para uma «humanização» em contexto educativo. **Educativ: boletim informativo**, p. 1-2,

fev./mar. 2014. Disponível em:

<https://repositorio.ipcb.pt/bitstream/10400.11/2404/1/A%20passagem%20da%20Web%20Henrique.pdf> . Acesso em: 19 abr. 2019.

GORTON, D. C. **Practical digital library generation into DSpace with the 5S framework.**

2007. Thesis (Master of Science IN Computer Science and Applications) - Virginia Tech, Virginia, USA, 2007. Disponível em:

https://vtechworks.lib.vt.edu/bitstream/handle/10919/31914/dgorton_thesis_final.pdf?sequence=1&isAllowed=y. Acesso em: 19 abr. 2019.

LAUFER, C. **Guia de Web Semântica.** Governo do Estado de São Paulo e Governo do Reino Unido. 2015. Disponível em:

<http://www.governoaberto.sp.gov.br/wp-content/uploads/2016/05/Book-Web-Semantica.pdf>. Acesso em: 15 abr. 2019.

MARCONDES, C. H. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 21, n. 2, p. 61-83, abr./jun. 2016. Disponível em:

<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2735/1748>. Acesso em: 15 abr. 2019.

MARTINS, D. L. *et al.* Repositório Digital com o Software Livre Tainacan: revisão da ferramenta e exemplo de implantação na área cultural com a Revista Filme Cultura. *In: XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 18., 2017, Marília. **Anais...** Marília: UNESP, 2017. Disponível em:

<http://enancib.marilia.unesp.br/index.php/xviiienancib/ENANCIB/paper/view/472>. Acesso em: 15 abr. 2019.

NOY, N. F.; MCGUINNESS, D. L. **Ontology Development 101: a guide to creating your first ontology.** 2001. Disponível em:

https://protege.stanford.edu/publications/ontology_development/ontology101.pdf . Acesso em: 15 abr. 2019.

OPEN ARCHIVES INITIATIVE. **Protocol for Metadata Harvesting.** 2019. Disponível em:

<https://www.openarchives.org/pmh/>. Acesso em: 15 abr. 2019.

PATRÍCIO, H. S. A Europeia e a agregação de metadados na web: análise dos esquemas ESE/EDM e da aplicação de standards da web semântica a dados de bibliotecas. **Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas**, n. 11, 2012. Disponível em:

<https://www.bad.pt/publicacoes/index.php/congressosbad/article/view/458/pdf>. Acesso em: 15 abr. 2019.

SANTARÉM SEGUNDO, José Eduardo. Web Semântica: introdução a recuperação de dados usando SPARQL. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 15., 2014, Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 2014. Disponível em:

http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/3191/2014_GT8-CO_09.pdf?sequence=1. Acesso em: 15 abr. 2019.

SPARQL BY EXAMPLE. **Tutorial.** 2018. Disponível em: <https://www.w3.org/2009/Talks/0615->

[qbe/](#). Acesso em: 15 abr. 2019.

W3SCHOOLS. **XML RDF**. 2018. Disponível em: https://www.w3schools.com/xml/xml_rdf.asp. Acesso em: 15 abr. 2019.

W3C RECOMMENDATION. **OWL Web Ontology Language Guide**. 2004. Disponível em: <https://www.w3.org/TR/2004/REC-owl-guide-20040210/>. Acesso em 15 abr 2019.

W3C RECOMMENDATION. **SPARQL Query Language for RDF**. 2008. Disponível em: <https://www.w3.org/TR/rdf-sparql-query/>. Acesso em: 15 abr. 2019.

W3C. **SIOC/WhatIsSIOC**. 2016. Disponível em: <https://www.w3.org/wiki/SIOC/WhatIsSIOC>. Acesso em: 15 abr. 2019.

WOOD, David et al. Linked Data. **Structured Data on the Web**. Shelter Island: Manning Publications Co, 2014.