

## Um protótipo Linked data para catalogação semântica de publicações

**Nilton Freitas Junior**

**Professor e coordenador de extensão e coordenador do curso de Sistemas de Informação do UNIFAMINAS – Muriaé/MG. Graduação em Sistemas de Informação pelo UNIFAMINAS, pós-graduação em Engenharia de Software pela Universidade Gama Filho, pós-graduação em Planejamento, Implementação e Gestão da EAD pela Universidade Federal Fluminense. Mestrando em Pesquisa Operacional e Inteligência Computacional pela Universidade Cândido Mendes, em Campos dos Goytacazes.**

**Mark Douglas de Azevedo Jacynto**

**Professor adjunto da coordenação de Ciência da Computação, da Universidade Cândido Mendes (UCAM), Campos dos Goytacazes - RJ, professor DIV da coordenação de Informática, do Instituto Federal Fluminense (IFF) - Campus Campos-Centro, Campos dos Goytacazes – RJ. Doutorado em Informática (2012), mestrado em Informática (2002) e graduação em Engenharia de Computação (1998), os três títulos obtidos no Departamento de Informática da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)**

**<http://dx.doi.org/10.1590/1981-5344/2664>**

*Este artigo apresenta a proposta e subsequente construção de um protótipo para catalogação semântica de publicações científicas. O protótipo desenvolvido segue os princípios Linked Data da nova Web Semântica proposta pelo consórcio W3C e encontra-se operacional. A proposta tem como diferencial o fato de todos os arquivos digitais das publicações serem anotados semanticamente com metadados semiestruturados inteligíveis por máquinas, permitindo que agentes de software nos auxiliem nas tarefas de busca, integração e processamentos de tais publicações, estabelecendo uma base de conhecimento Linked Data de publicações. A pesquisa que ampara o estudo e desenvolvimento do protótipo apresenta conceitos inerentes à Web Semântica, ao padrão de dados semiestruturados Resource*

*Description Framework (RDF) e o uso de ontologias para representação de domínios de conhecimento. Testes de funcionamento são demonstrados e foram realizados com algumas publicações também relacionadas ao tema, salientando a capacidade do protótipo em fornecer informações acessíveis a pessoas e a máquinas (agentes de software), por meio de negociação de conteúdo. O protótipo encontra-se hospedado e está disponível para verificação online por qualquer interessado no tema deste artigo.*

**Palavras-chave:** *Catalogação semântica de publicações, Linked Data, Web Semântica, RDF.*

## **A Linked Data prototype for semantic cataloguing of publications**

*This article presents the proposal and subsequent construction of a prototype for semantic cataloguing of scientific publications. The prototype developed follows the Linked Data principles of the novel Semantic Web proposed by the W3C consortium and is operational. The proposal has as differential the fact all publications' digital files be semantically annotated with semi-structured metadata intelligible by machines, allowing software agents assist us in the tasks of searching, integration and processing of such publications, establishing a Linked Data knowledge base of publications. The research that supports the study and development prototype presents concepts inherent to the Semantic Web, the standard of semi-structured data Resource Description Framework (RDF) and the use of ontologies for representing knowledge domains. Functional tests are shown and were engaged with some publications also related to the theme, demonstrating the prototype's ability to provide information accessible to people and machines (software agents), through content negotiation. The prototype is hosted and available for online checking for any interested in the subject of this article.*

**Keywords:** *Semantic Cataloguing of Publications, Linked Data, Semantic Web, RDF.*

Recebido em 11.12.2015 Aceito em 30.08.2016

## 1 Introdução

Considerar sobre o avanço da Tecnologia da Informação (TI) na sociedade moderna é uma retórica comumente encontrada em inúmeros estudos sobre este assunto. Sim, a TI mantém sólido e constante avanço em todas as suas vertentes de aplicação, seja na infraestrutura de hardware, na ampliação dos meios de comunicação, nos diversos softwares produzidos e muito mais. Este fato suscita uma questão importante para as pesquisas que envolvem a TI: a necessidade de se produzir padrões e técnicas capazes de apoiar este crescimento.

É inegável que a Web figura contemporaneamente como um dos pontos mais centrais no crescimento da TI, pois requer e estimula desenvolvimentos paralelos diversos. Sobre um desenvolvimento em particular está o tema deste trabalho, que é a catalogação de publicações (artigos científicos, teses e monografias, entre outros) como recursos digitais anotados semanticamente em consonância com os padrões e tecnologias da Web Semântica propostos pelo consórcio W3C<sup>1</sup>. Em outras palavras, recursos descritos explicitamente por metadados estruturados inteligíveis por máquina, recursos estes que são instâncias de modelos de representação de conhecimento compartilhados chamados ontologias, capazes de especificar conceitos e seus relacionamentos, tornando, portanto, explícito o significado destes recursos para a máquina.

A arquitetura da Web Semântica é alicerçada sobre um modelo de metadados padrão chamado *Resource Description Framework* (RDF) (CYGANIAK, WOOD e LANTHALER, 2014), definindo uma organização lógica em termos de estrutura para apoiar representações, acessos, restrições e relacionamentos entre recursos diversos, mesmo que estes não habitem o mesmo ambiente de desenvolvimento (CURÉ; BLIN, 2015, p. 43). Assim como foi feito no início da história da rede mundial de computadores, com protocolos unindo diferentes plataformas de trabalho, o padrão RDF pode ligar dados com origens diversas, dando forma ao novo paradigma da Web Semântica.

Este cenário tende a promover a construção de novos *softwares* para gerenciamento de conteúdo digital que ofereçam a publicação de dados em formato RDF. O objetivo deste trabalho é apresentar a construção de um protótipo capaz de gerenciar a catalogação de publicações científicas com recursos de usabilidade tradicionais voltados para humanos, bem como disponibilização de dados estruturados em RDF voltados para máquinas (agentes de software), contribuindo para a construção da nova versão da Web, chamada de Web Semântica ou Web de Dados Ligados ou, ainda, simplesmente Web 3.0.

O restante do artigo está organizado da seguinte forma: a próxima seção descreve, brevemente, os conceitos sobre o modelo RDF e os princípios *Linked Data*; a seção seguinte apresenta a definição de ontologias como modelo de representação formal do conhecimento; o desenvolvimento e apresentação de algumas imagens de funcionamento

---

<sup>1</sup> Disponível em: <<http://www.w3.org>>. Acesso em: 21 abr. 2015.

do protótipo são descritos na seção subsequente; a seguir, a seção sobre os resultados e discussões e, por fim, a seção que aborda as considerações finais, bem como alguns trabalhos futuros.

## 2 RDF e a concepção de dados ligados

Em fevereiro de 2009, Tim Berners-Lee apresentou uma palestra<sup>2</sup> na qual discorreu sobre seu trabalho com a criação “de um sistema global de hipertexto” (BERNERS-LEE, 2009). Sua motivação principal era publicar e disponibilizar acesso a diversos documentos de pesquisa, independente de sua plataforma de criação. Assim nascia a *Web* como conhecemos.

Durante sua apresentação, Berners-Lee conclama a publicar não só documentos, mas sim dados sobre estes documentos, dados que estejam ligados a outros dados e que permitam uma estruturação do conteúdo disponibilizado na *Web*. Heath e Bizer (2011, p. 1, tradução nossa) lembram que “nós estamos cercados por dados, seja sobre desempenho de nossas escolas, sobre a eficiência dos combustíveis em nossos carros e sobre a multiplicidade de produtos de diferentes fornecedores”. De fato, são dados processados que geram informação, e isso não acontece exclusivamente na *Web*.

Não obstante, dados distribuídos na *Web* requerem um mecanismo padrão para especificar suas instâncias e proporcionar meios de interligação com outros dados, usando a infraestrutura pré-existente da *Web* convencional. Este mecanismo é o *Resource Description Framework*, ou simplesmente RDF, um modelo de dados em grafo que permite estabelecer declarações de relacionamentos entre quaisquer “coisas” publicadas na *Web* e que possam ser descritas (HEATH; BIZER, 2011, p. 4).

Um grafo RDF é um conjunto de triplas recurso-propriedade-valor, usadas para descrever recursos, onde um recurso é qualquer “coisa” do mundo real identificada por um endereço *Web*. O uso de *Uniform Resource Identifier* (URI) garante a identificação única de um recurso e como este pode ser acessado (BORGHOFF *et al.*, 2003, p. 92). Uma propriedade também é um URI que identifica um elemento de significado, definido por uma ontologia reconhecida. Valores assumem literais tradicionalmente conhecidos ou também podem ser outro URI, perfazendo relacionamento entre recursos publicados na *Web*. A Figura 1 mostra um exemplo de tripla RDF descrevendo uma relação entre o recurso *Machado de Assis*, identificado pelo respectivo URI, associado a um valor que, neste exemplo, também é um URI, referente ao livro *Dom Casmurro*. A relação é estabelecida usando a propriedade “*author*”, de uma ontologia, que também é identificada por seu próprio URI.

---

<sup>2</sup> Disponível em <[https://www.ted.com/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_web](https://www.ted.com/talks/tim_berniers_lee_on_the_next_web)>. Acesso em: 1º nov. 2014.

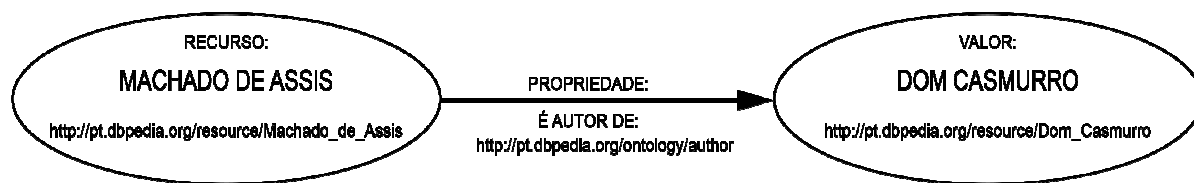


Figura 1 - Exemplo de um conjunto de tripla RDF

Fonte: Os autores.

Como subconjunto mais pragmático da *Web Semântica*, em Bernes-Lee (2009), é descrito o conceito de *Linked Data*, quatro princípios para publicação e interligação de dados RDF, visando construir a *Web de Dados Ligados* – um espaço global de dados estruturados, totalmente compreensíveis por agentes de software. São eles:

- a) usar URIs como nomes para as "coisas" (recursos) disponibilizadas na *Web*;
- b) usar HTTP URIs para que as pessoas ou máquinas possam procurar e acessar esses nomes, usando o protocolo HTTP da *Web* convencional;
- c) quando alguém desreferenciar (acessar) um URI de um recurso, esta deve fornecer informações úteis em RDF e seus padrões, descrevendo explicitamente o recurso;
- d) incluir *links* para outros URIs, de modo que as pessoas ou máquinas possam navegar por estes *links* e descobrir mais "coisas" (recursos).

Esta arquitetura é preferencialmente voltada para outros sistemas, a fim de executar processamento de dados e estabelecer novos relacionamentos, como sugere a concepção de dados abertos ligados (*Open Linked Data*). Tal contexto é muito bem apresentado por Marcondes (2012), que explica:

A proposta de dados abertos interligados oferece grande potencial ao conectar recursos informacionais através de *links* semânticos, *links* que são significativos também para programas. Ao contrário, *links* convencionais nada mais são (além de uma eventual etiqueta textual significativa para usuários humanos) que meios para que programas navegadores, a partir de um recurso, acessem outro, sem explicitar qual o significado da ligação entre os recursos. Sendo significativos para programas, *links* semânticos podem ser processados de forma mais rica por eles, explorando e enriquecendo cognitivamente o significado (legível por

máquina) da ligação entre ambos os recursos (MARCONDES, 2012, p. 3).

Enquanto na *Web* tradicional temos a navegação entre documentos estabelecida apenas por *links* diretos, apenas para um único novo recurso, na *Web* de Dados Ligados, um *link* pode fornecer tantos dados quanto forem possíveis para identificar e associar semântica a qualquer "coisa" que possa ser disponibilizada como um recurso.

É possível que o melhor exemplo contemporâneo de dados ligados seja o projeto *DBpedia*<sup>3</sup>, que desenvolve "um esforço comunitário com a finalidade de extrair informações estruturadas da *Wikipedia*<sup>4</sup> e tornar essa informação disponível na *Web* em RDF" (*DBpedia*, 2015, tradução nossa). Por isso, a fonte de dados *DBpedia* foi escolhida para ser utilizada no protótipo deste trabalho, para perfazer o *Mashup Linked Data*, enriquecendo, sobremaneira, a descrição RDF das publicações com informação oriunda da *Wikipedia*.

Especificamente no contexto deste trabalho, uma publicação armazenada e gerenciada pelo protótipo desenvolvido, ao ser acessada por meio do seu URI, oferece mais do que apenas sua representação via página HTML convencional, mas, sobretudo, sua representação via arquivo RDF compreensível por máquina, reusando propriedades de ontologias consagradas e interligando com recursos da *DBpedia* relevantes que agregam valor à informação retornada.

### 3 Ontologias e a Web Semântica

Segundo Gruber (1993)<sup>5</sup> *apud* Breitman (2005, p. 30), a definição de ontologia relacionada à *Web* Semântica, mais frequentemente encontrada, é uma especificação formal e explícita de uma conceitualização compartilhada. Em outras palavras, isso implica duas necessidades básicas: uma "coisa" tem que estar na *Web* e tem que oferecer significado explícito para máquinas. Sem isso, não se tem a *Web* Semântica.

Uma ontologia é criada de forma a associar nomes a relações entre entidades e aquilo que representam em um universo definido. Por isso, ontologias também são conhecidas por vocabulários, significando que a escolha de uma ontologia deve ser adequada ao objetivo que se espera alcançar na questão de atribuir significados e o reconhecimento dos mesmos por sistemas que irão interpretar as triplas RDF produzidas por uma aplicação.

Uma analogia interessante pode ser feita sobre a relevância das inúmeras ontologias hoje existentes, observando-se os princípios descritos na obra de Charles Darwin, *A origem das espécies*:

<sup>3</sup> Disponível em: <<http://dbpedia.org>>. Acesso em: 21 abr. 2015.

<sup>4</sup> Wikipédia é um projeto de enciclopédia multilíngue, de licença livre, baseado na *Web*, desenvolvido de maneira colaborativa. Disponível em: <<http://pt.wikipedia.org/wiki/Wikip%C3%A9dia>>. Acesso em: 21 abr. 2015

<sup>5</sup> GRUBER, T. R. *et al.* A translation approach to portable ontology specifications. *Knowledge acquisition*, v. 5, n. 2, p. 199-220, 1993.

Metaforicamente pode-se dizer que a seleção natural procura a cada dia momento, em todo lugar, as mais tênues variações, rejeitando as nocivas, conservando e ampliando todas as que forem úteis, trabalhando silenciosa e imperceptivelmente, quando e onde quer que se ofereça a oportunidade, pelo aperfeiçoamento de cada ser vivo com relação a suas condições de vida orgânicas e inorgânicas (DARWIN, 2009, p. 81).

À medida que as ontologias são compartilhadas por um número maior de sistemas que gozem de reconhecimento e reputação dentro do universo da *Web*, mais consistentes elas serão, recebendo maior aperfeiçoamento. Não necessariamente uma ontologia deve desaparecer se não for utilizada em larga escala, mas o crescimento direcionado de uma determinada ontologia evita redundâncias de significados. Resumindo, ontologias devem ser reutilizadas sempre que possível para um dado domínio de conhecimento.

#### 4 Desenvolvimento do protótipo de aplicação semântica

Ao se trabalhar com a Tecnologia da Informação para organizar a própria informação, busca-se tanto os processos de tratamento descritivo quanto os tratamentos temáticos, ou seja, os conteúdos dos documentos. Os resultados de qualquer desenvolvimento neste sentido precisam buscar a identificação do conteúdo dos documentos por meio da busca e evidência dos principais assuntos dentro do próprio documento (REZENDE, 2005, p. 352).

Tal pensamento ajuda justificar a pesquisa e o desenvolvimento de aplicações que contemplem as necessidades descritas e que também sejam funcionais, oferecendo real possibilidade de utilização. Por isso, no desenvolvimento do protótipo foi utilizada a linguagem de programação PHP<sup>6</sup>, que é de código livre, possuindo recursos adequados para a construção da aplicação, além de permitir que esta seja utilizada tanto em plataforma *desktop*, quanto para plataforma *Web*.

O protótipo conta também com o uso de banco de dados relacional MySQL<sup>7</sup>, simulando um banco de dados RDF (*triple store*) por meio da biblioteca RDF API for PHP (RAP)<sup>8</sup>. A parceria entre PHP e MySQL é encontrada facilmente nos serviços de hospedagem mais comerciais, o que foi levado em consideração nesta escolha. A interface gráfica foi implementada com emprego da linguagem HTML5<sup>9</sup>, amparada na utilização do *framework Bootstrap*<sup>10</sup>. Para a representação RDF de

<sup>6</sup> Disponível em: <<http://php.net/>>. Acesso em: 20 nov. 2014.

<sup>7</sup> Disponível em: <<https://www.mysql.com/>>. Acesso em: 20 nov. 2014.

<sup>8</sup> Disponível em: <<http://wifo5-03.informatik.uni-mannheim.de/bizer/rdffapi/index.html>>. Acesso em: 20 nov. 2014.

<sup>9</sup> Disponível em: <<http://www.w3.org/TR/html5/>>. Acesso em: 20 nov. 2014.

<sup>10</sup> *Bootstrap* é um conjunto *open source* de padrões para os elementos HTML, CSS e Javascript mais utilizados, agilizando o desenvolvimento de interface (*front-end*) para páginas Web, podendo ser também personalizado conforme as necessidades do desenvolvimento.

publicações e registros relacionados, foram selecionadas, no portal *Linked Open Vocabularies (LOV)*<sup>11</sup>, as seguintes ontologias *Linked Data*:

*The RDF Schema vocabulary (RDFS)*: *RDF Schema* (BRICKLEY; GUHA, 2014) é uma extensão do vocabulário básico RDF, e fornece um vocabulário de modelagem de dados (classe e propriedades) para dados RDF.

*DCMI Metadata Terms (DCTERMS)*: ontologia usada para descrever recursos em geral. Uma especificação atualizada de todos os termos de metadados mantida pela *Dublin Core Metadata Initiative (DUBLIN CORE METADATA INITIATIVE, 2014)*, incluindo propriedades, esquemas de codificação de vocabulário, esquemas de codificação de sintaxes e classes.

*Friend of a Friend vocabulary (FOAF)*: ontologia usada para descrever pessoas. FOAF (BRICKLEY; MILLER, 2014) é um projeto dedicado a unir as pessoas e informações usando a *Web*. Independentemente de a informação estar na cabeça das pessoas, em documentos físicos ou digitais, ou sob a forma de dados factuais, ela pode ser ligada.

*Simple Knowledge Organization System (SKOS)*: ontologia usada para criação de taxonomias e tesouros. O Sistema Simples de Organização do Conhecimento (em inglês, SKOS) (MILES; BECHHOFER, 2009) é um modelo de dados comum para compartilhamento e ligação dos sistemas de organização do conhecimento através da *Web Semântica*.

Os recursos do vocabulário RDF são utilizados para identificar tipos (classes) nos cadastros de autores, palavras-chave, tipos de publicações e publicações. A ontologia DCTERMS compõe os registros das publicações, fazendo jus ao objetivo da iniciativa dos metadados *Dublin Core (DUBLIN CORE METADATA INITIATIVE, 2012)*, para que sejam promovidos seus padrões de descrição de recursos na *Web*. Autores são descritos pela ontologia FOAF e, finalmente, as palavras-chave e tipificações das publicações cadastradas são descritas através da ontologia SKOS, utilizada neste caso também como um vocabulário padronizado (folksonomia semântica).

O protótipo da aplicação dispõe de uma sequência de cadastros necessários para que uma publicação seja efetivamente inserida em sua base de dados RDF. Para cada entidade (palavra-chave, autor, tipo de publicação e publicação) cadastrada é gerada automaticamente seu respectivo URI, permitindo que o grafo RDF subjacente ao cadastro seja criado. Os URIs criados são desreferenciáveis, permitindo visualização dos registros tanto em HTML, para pessoas, quanto em RDF para outros

<sup>11</sup> Disponível em: <<http://lov.okfn.org/dataset/lov/>>. Acesso em: 20 nov. 2014



sistemas (agentes de software). Um menu principal de opções oferece acessos a estes cadastros. Nesta versão do protótipo, é necessário que sejam cadastradas as palavras-chave (Figura 2) e os autores (Figura 3), necessariamente nesta ordem, pois cada autor inserido no sistema poderá ter associadas a ele várias palavras-chave como tópicos de interesse.

**1 Cadastro de Palavras-chave**

Editar registro

Palavra-chave: Machado de Assis

Definição: Importante escritor brasileiro, considerado maior nome da literatura brasileira.

**OFEREÇA MAIS DETALHES PARA ESTA PALAVRA-CHAVE:**

Nome alternativo: Joaquim Maria Machado de Assis

Conceito mais amplo: Escritor X

Conceito Relacionado: Literatura X

Permitir busca na Dbpedia:  SIM  NÃO

Figura 2 - Cadastro de palavras-chave

Fonte: Autoria própria a partir dos dados da publicação de Santos Neto *et al.* (2013).

A Figura 2 demonstra o registro de uma das palavras-chave encontradas na publicação de Santos Neto *et al.* (2013), ilustrando o registro que é feito para todos os termos utilizados como palavras-chave em uma publicação. O registro também está ligado a um recurso externo da *DBpedia* ([http://dbpedia.org/resource/Machado\\_de\\_Assis](http://dbpedia.org/resource/Machado_de_Assis))<sup>12</sup>, que pode ser utilizado para possíveis *mashups* de conteúdo, em associação à publicação que possui esta palavra-chave ou qualquer outro relacionamento que venha a ser estabelecido dentro da aplicação, situação que pode ser percebida na Figura 3, na qual os dados de um dos autores em Santos Neto *et al.* (2013) contam também com seus “interesses” dentro da aplicação; não por acaso, as palavras-chave utilizadas na publicação da qual participa, mas podendo utilizar qualquer outro registro já existente no banco de dados do protótipo.

<sup>12</sup> Machado de Assis. Recurso da DBpedia. Disponível em <[http://dbpedia.org/page/Machado\\_de\\_Assis](http://dbpedia.org/page/Machado_de_Assis)>. Acesso em: 1º mar. 2015.

Figura 3 - Cadastro de autores

Fonte: Autoria própria a partir dos dados da publicação de Santos Neto *et al.* (2013).

Uma vez disponíveis no sistema as palavras-chave, autores e também os tipos de publicação, faz-se a inclusão de uma publicação propriamente dita (Figura 4).

Figura 4 - Cadastro de uma publicação

Fonte: Autoria própria a partir dos dados da publicação de Santos Neto *et al.* (2013).

Conforme dito anteriormente, para todos estes cadastros, são gerados, internamente no sistema, os respectivos subgrafos RDF. A Tabela 1 apresenta as triplas RDF subjacentes aos cadastros da publicação (Figura 4) e da palavra-chave (Figura 2). Estes dados podem ser obtidos desreferenciando o URI <http://www.niltonfjunior.com.br/pws/resource/publication/6>, para a publicação, e o URI [http://www.niltonfjunior.com.br/pws/resource/keyword/Machado\\_de\\_Assi](http://www.niltonfjunior.com.br/pws/resource/keyword/Machado_de_Assi)

s, para a palavra-chave. Para abreviar os URIs da Tabela 1, serão considerados os seguintes prefixos para os namespaces envolvidos:

prefixo dbpedia: <http://dbpedia.org/resource/>  
 prefixo dc: <http://purl.org/dc/elements/1.1/>  
 prefixo dcmitype: <http://purl.org/dc/dcmitype/>  
 prefixo dcterms: <http://purl.org/dc/terms/>  
 prefixo ns1: <http://www.niltonfjunior.com.br/pws/resource/>  
 prefixo rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>  
 prefixo rdfs: <http://www.w3.org/2000/01/rdf-schema#>  
 prefixo skos: <http://www.w3.org/2004/02/skos/core#>

Tabela 1 - triplas subjacentes aos cadastros

RECURSO	PROPRIEDADE	VALOR
ns1:publication/6	rdf:type	dcterms:BibliographicResource
ns1:publication/6	rdf:type	dcmitype:Text
ns1:publication/6	dc:type	ns1:type/artigo
ns1:publication/6	dc:title	"Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: um caso machadiano"@PT
ns1:publication/6	dc:creator	ns1:people/carlosmarcondes5
ns1:publication/6	dc:creator	ns1:people/antoniosantosneto4
ns1:publication/6	dcterms:abstract	"O artigo propoe um caso ficticio para ilustrar o uso das tecnologias de Dados Abertos Interligados como mecanismos para tornar interoperaveis informacoes em acervos de bibliotecas, arquivos e museus, utilizando como exemplo as informacoes sobre o escritor Machado de Assis e sua obra. Utilizou-se como metodo o estudo de um caso restrito e ilustrativo, nao exaustivo. Aplicaram-se diferentes vocabularios para interligacao dos dados e para sua representacao, utilizando os principios do Linked Data; sugere-se que os profissionais da informacao usufruam dessa potencialidade e busquem novas aplicacoes para ampliar a interoperabilidade dos dados disponiveis na Web."@PT
ns1:publication/6	dc:subject	ns1:keyword/Web_Semântica
ns1:publication/6	dc:subject	ns1:keyword/Machado_de_Assis
ns1:publication/6	dc:subject	ns1:keyword/Ontologias
ns1:publication/6	dc:subject	ns1:keyword/Dados_Abertos_Interligados
ns1:publication/6	dc:subject	ns1:keyword/Resource_Description_Framework
ns1:publication/6	dcterms:hasVersion	ns1:pdf/6

ns1:keyword/Machado_de_Assis	skos:related	ns1:keyword/Literatura_Brasileira
ns1:keyword/Machado_de_Assis	skos:altLabel	"Joaquim Maria Machado de Assis"@PT
ns1:keyword/Machado_de_Assis	skos:prefLabel	"Machado de Assis"@PT
ns1:keyword/Machado_de_Assis	rdfs:seeAlso	dbpedia:Machado_de_Assis
ns1:keyword/Machado_de_Assis	rdf:type	skos:Concept
ns1:keyword/Machado_de_Assis	skos:definition	"Importante escritor brasileiro, considerado maior nome da literatura brasileira."@PT
ns1:keyword/Machado_de_Assis	skos:broader	ns1:keyword/Escritor

Fonte: Os autores.

Vale destacar na Tabela 1 o uso da propriedade *rdfs:seeAlso* (em português, *veja também*) do vocabulário RDFS, utilizada para perfazer o *Mashup Linked Data* com a fonte de dados *DBpedia*. No caso, temos a interligação da palavra-chave "Machado de Assis" do sistema com o recurso "Machado de Assis" da *DBpedia* ([http://dbpedia.org/resource/Machado\\_de\\_Assis](http://dbpedia.org/resource/Machado_de_Assis))<sup>13</sup>, permitindo que mais informações acerca de *Machado de Assis* sejam extraídas (reusadas) desta valiosa fonte *Linked Data*, enriquecendo ainda mais a base de conhecimento do sistema.

Para experimentar o funcionamento do protótipo, o mesmo encontra-se disponibilizado *online*, no endereço <http://www.niltonfjunior.com.br/pws/>. O acesso requer informação de *login* e senha, ambos definidos como *admin*. A disponibilização do protótipo será mantida por período indeterminado, oferecendo aos interessados e estudiosos do assunto a oportunidade de verificar na prática o contexto descrito neste trabalho.

## Resultados e discussões

Em sua primeira versão disponibilizada, o protótipo da aplicação semântica apresenta funcionamento adequado à proposta de realizar cadastros referentes a publicações científicas. A implantação dos códigos-fonte em um serviço de hospedagem contratado transcorreu dentro da normalidade requerida neste tipo de aplicação.

Para ampliar os testes com publicações reais, foi feita uma pesquisa no sistema *Scielo*, na qual foram encontrados oito resultados para o termo "Web semântica", listados a seguir.

- 1 AMBINDER, Déborah Motta et al. Novas experiências para apresentação, acesso e leitura de artigos científicos digitais na web. *TransInformação*, Campinas, v. 25, n. 3, p. 195-201, set./dez., 2013.;

<sup>13</sup> Machado de Assis. Recurso da *DBpedia*. Disponível em <[http://dbpedia.org/page/Machado\\_de\\_Assis](http://dbpedia.org/page/Machado_de_Assis)>. Acesso em: 1º mar. 2015.

- 2 LARA, Marilda Lopes Ginez de. Documentary languages and knowledge organization systems in the context of the semantic web. *Transinformação*, v. 25, n. 2, p. 145-150, 2013.;
- 3 SANTOS NETO, Antonio Laurindo dos *et al.* Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: um caso machadiano Using open data technology to connect libraries, archives and museums: a Machadian case. *Transinformação*, v. 25, n. 1, p. 81-87, 2013.;
- 4 CATARINO, Maria Elisabete; DE SOUZA, Terezinha Batista. A representação descritiva no contexto da web semântica. *Transinformação*, v. 24, n. 2, 2012.;
- 5 PICKLER, Maria Elisa Valentim. Web Semântica: ontologias como ferramentas de representação do conhecimento. *Perspectivas em Ciência da Informação*, v. 12, n. 1, p. 65-83, 2007.;
- 6 CASTRO, F. F. de; SANTOS, P. L. V. A. da C. Uso das tecnologias na representação descritiva: o padrão de descrição bibliográfica semântica MarcOnt Initiative nos ambientes informacionais digitais. *Ci. Inf. [online]*, v. 38, n. 1, p. 74-85. 2009. ISSN 1518-8353.;
- 7 SOUZA, R. R.; ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. *Ci. Inf. [online]*, v. 33, n. 1, p. 132-141, 2004. ISSN 1518-8353.;
- 8 PERISSE, M. C.; NARVAEZ, J. L. La web semántica en la educación superior. *JISTEM - J.Inf.Syst. Technol. Manag. [online]*, v. 5, n. 2, p. 223-234, 2008. ISSN 1807-1775.

O exemplo de utilização do protótipo apresentado anteriormente, por meio das Figuras 2, 3 e 4 foi construído a partir dos dados de uma dessas publicações: Santos Neto *et al.* (2013).

À primeira vista, a interface do protótipo lembra um sistema de informação tradicional. Pode-se dizer que isto é esperado, porque os princípios *Linked Data* (BERNERS-LEE, 2009) não focam em diferenças relacionadas à interface de uma aplicação, mas sim o quanto ela está preparada para oferecer dados identificados por URIs que sejam padronizados em um formato reconhecido por outros sistemas e que estejam, obviamente, ligados a outros dados.

Manter uma interface tradicional pode ser um fator de estímulo ao uso de aplicações semânticas, visto que um usuário final deste tipo de sistema não deve se importar em como os dados são gerados, mas sim receber os benefícios semânticos oferecidos pelo novo paradigma. Os dados gerados, embora acessíveis a qualquer usuário da aplicação, estão preparados para leitura de outros sistemas através de redirecionamento

automático do URI de cada recurso, quando a aplicação identifica este tipo de solicitação.

A participação do usuário deve ser voltada para modelos de *folksonomia*, ou seja, classificar os recursos por sua própria percepção. Por meio de mecanismos de mineração de texto, o sistema oferece possibilidade para catalogação de novos termos como palavras-chave de uma publicação.

O protótipo conta com uma experiência inicial de dados ligados à *DBpedia* por meio do cadastro de palavras-chave, perfazendo o *Mashup Linked Data*. Cada registro inserido pode ser usado como parâmetro para uma consulta SPARQL<sup>14</sup> (HARRIS; SEABORNE, 2013) feita diretamente na *DBpedia*. A consulta mostra-se perfeitamente funcional, mas testes demonstraram que são raros os retornos de dados em língua portuguesa.

Obviamente, a pequena quantidade de dados ligados em língua portuguesa disponíveis na *DBpedia* tende a aumentar à medida que outros sistemas baseados em RDF forem desenvolvidos, embora o protótipo também possa oferecer buscas internas em seus dados muito mais precisas, por meio de consultas SPARQL.

Com o devido desenvolvimento, um sistema baseado em RDF pode produzir informações mais completas, por meio de resultados obtidos por consultas SPARQL federadas, ou seja, consultas envolvendo várias fontes de dados da *Web* de Dados Ligados. A geração de dados em formato RDF elimina fronteiras de sistema e promove a ideia de mundo aberto para seus dados.

## Considerações finais

Este trabalho apresentou o desenvolvimento de um protótipo de aplicação baseado nos princípios *Linked Data* da *Web Semântica*, os quais preveem a utilização de URIs para descrever recursos na *Web* e a estruturação dos dados em formato padrão RDF. O protótipo finalizado encontra-se disponibilizado para verificação de suas funcionalidades.

Não há pretensões de classificar o protótipo apresentado como uma aplicação totalmente completa, no que diz respeito aos princípios *Linked Data* da *Web Semântica*. Entretanto, pode-se afirmar que os objetivos traçados pelo trabalho foram alcançados quando se observa que:

- 1 A aplicação provê URIs como nomes para seus recursos;
- 2 Uma vez estando disponibilizada em um serviço de hospedagem *Web*, clientes (pessoas ou sistemas) podem procurar pelos recursos diretamente por seus URIs;
- 3 Todos os URIs de recursos disponíveis na aplicação são desreferenciáveis, tanto para pessoas (HTML) quanto para outros sistemas (RDF);

---

<sup>14</sup> SPARQL é a linguagem de consulta padrão para o modelo de dados RDF. SPARQL está para RDF, assim como SQL está para o modelo relacional.

4 Os dados estão estruturados em formato RDF, disponibilizados sem restrições, oferecendo também links semânticos para outros recursos da *Web* de Dados.

Desenvolvimentos mais aprofundados para o protótipo serão bem-vindos, seja nas bibliotecas utilizadas ou mesmo em recursos de usabilidade. A boa disponibilidade de serviços de hospedagem *Web* que oferecem a plataforma PHP/MySQL pode promover tal desenvolvimento, uma vez que o modelo de dados RDF torne-se comercialmente mais forte.

De fato, o desenvolvimento de sistemas gerenciadores de bancos de dados apropriados para a *Web* Semântica deve se preocupar com o crescimento deste modelo de aplicação em seus detalhes. Parafraseando Jacyntho (2012, p. 17), a maioria das iniciativas para bancos de dados em triplas tem abordado apenas consultas nestes repositórios, mas há muitas indagações sobre o que aconteceria com a escalabilidade de transações nestas plataformas, o que claramente tende a acontecer.

Como trabalhos futuros, pretende-se ampliar o uso dos dados ligados à *DBpedia* (e outras fontes de dados) além das palavras-chave e, a partir destes *links* externos à aplicação, promover *mashup* de conteúdo. Também há a possibilidade de implantação do protótipo em ambiente real de trabalho, controlado, no qual poderão ser levantados dados de sua utilização e aceitação por parte de seus usuários, a fim de verificar seu desempenho e promover suas funcionalidades dentro da *Web* Semântica. Outro ponto importante que precisa ser melhor explorado são as consultas SPARQL com inferência ontológica, onde as máquinas com base nos axiomas das ontologias empregadas podem, automaticamente, deduzir novas triplas RDF, aumentando o resultado retornado pelas consultas, trazendo à tona valioso conhecimento para nos auxiliar nas buscas pelas publicações.

## Referências

AMBINDER, D. M.; MARCONDES, C. H. Novas experiências para apresentação, acesso e leitura de artigos científicos digitais na *Web*. *Transinformação*, Campinas, v. 25, n. 3, dez. 2013. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-37862013000300002&lng=pt&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862013000300002&lng=pt&nrm=iso)>. Acesso em: 1º mar. 2015.

BERNERS-LEE, T. Linked data. *Design Issues*, 2009. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 21 dez. 2014.

BERNERS-LEE, T. *On the next Web*. TED 2009. Filmed Feb. 2009. Disponível em: <[http://www.ted.com/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_Web?language=pt-br](http://www.ted.com/talks/tim_berniers_lee_on_the_next_Web?language=pt-br)>. Acesso em: 25 abr. 2014.

BORGHOFF, U. M. *et al. Long-term preservation of digital documents: principles and practices*. California: Springer, 2003.

BREITMAN, K. K. *Web semântica: a internet do futuro*. Rio de Janeiro: LTC, 2005.

BRICKLEY, D.; GUHA, R. V. *RDF Schema 1.1*. 2014. Disponível em: <<http://www.w3.org/TR/rdf-schema/>>. Acesso em: 21 abr. 2015.

BRICKLEY, D.; MILLER, L. *FOAF Vocabulary Specification 0.99*. 2014. Disponível em: <<http://xmlns.com/foaf/spec/>>. Acesso em: 21 abr. 2015.

CASTRO, F. F. de; SANTOS, P. L. V. A. da C. Uso das tecnologias na representação descritiva: o padrão de descrição bibliográfica semântica MarcOnt Initiative nos ambientes informacionais digitais. *Ci. Inf. [online]*, v. 38, n. 1, p. 74-85. 2009. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/view/1082>>. Acesso em: 16 ago. 2015.

CATARINO, M. E.; SOUZA, T. B. de. A representação descritiva no contexto da *Web semântica*. *Transinformação*, Campinas, v. 24, n. 2, ago. 2012. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-37862012000200001&lng=pt&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862012000200001&lng=pt&nrm=iso)>. Acesso em: 1º mar. 2015.

CURÉ, Olivier; BLIN, Guillaume. *RDF Database Systems: Triples Storage and SPARQL Query Processing*. Morgan Kaufmann, 2014. Disponível em: <<https://books.google.com.br/books?isbn=0128004703>>. Acesso em: 06 fev. 2015.

CYGANIAK, R.; WOOD, D.; LANTHALER, M. *RDF 1.1: concepts and abstract syntax*. 2014. Disponível em: <<http://www.w3.org/TR/rdf11-concepts/>>. Acesso em: 21 abr. 2015.

DARWIN, C. *A origem das espécies por meio da seleção natural ou a preservação das raças favorecidas na luta pela vida*. Tradução. São Paulo: Escala, 2009.

DUBLIN CORE METADATA INITIATIVE. Usage Board. *DCMI Metadata Terms*. 2012. Disponível em: <<http://dublincore.org/documents/dcmi-terms/>>. Acesso em: 11 maio 2015.

DUBLIN CORE METADATA INITIATIVE. The Metadata community: supporting innovation in metadata design, implementation & best practices. 2014. Disponível em: <<http://dublincore.org/>>. Acesso em: 6 dez. 2014.

HARRIS, S.; SEABORNE, A. *SPARQL 1.1 query language: W3C recommendation*. 21 March 2013. Disponível em: <<http://www.w3.org/TR/sparql11-query/>>. Acesso em: 11 maio 2015.

HEATH, Tom; BIZER, Christian. Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, v. 1, n. 1, p. 1-136, 2011. Ebook disponível em: <<http://linkeddatatbook.com/editions/1.0/>>. Acesso em: 06 fev. 2015.



JACYNTHO, M. D. de A.. *Um modelo de bloqueio multigranular para RDF*. Tese (Doutorado em Informática) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, Rio de Janeiro, 2012. Disponível em: <[http://www.maxwell.vrac.puc-rio.br/Busca\\_etds.php?strSecao=resultado&nrSeq=20236@2](http://www.maxwell.vrac.puc-rio.br/Busca_etds.php?strSecao=resultado&nrSeq=20236@2)>. Acesso em: 20 abr. 2015.

LARA, Marilda Lopes Ginez de. Documentary languages and knowledge organization systems in the context of the semantic web. *Transinformação*, v. 25, n. 2, p. 145-150, 2013. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-37862013000200005&lng=pt&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862013000200005&lng=pt&nrm=iso)>. Acesso em: 1º mar. 2015.

MARCONDES, C. H. "Linked data": dados interligados e interoperabilidade entre arquivos, bibliotecas e museus na Web. *Revista Eletrônica de Biblioteconomia e Ciência da Informação*, v. 17, n. 34, p. 171-192, maio./ago. 2012. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17n34p171>>. Acesso em: 1º mar. 2015.

MILES, A.; BECHHOFFER, S. *SKOS Simple Knowledge Organization System Namespace Document*. 2009. Disponível em: <<http://www.w3.org/2009/08/skos-reference/skos.html>>. Acesso em: 6 dez. 2014.

PERISSE, M. C.; NARVAEZ, J. L. La web semántica en la educación superior. *JISTEM - J.Inf.Syst. Technol. Manag. [online]*, v. 5, n. 2, p. 223-234, 2008. Disponível em: <<http://www.scielo.br/pdf/jistm/v5n2/03.pdf>>. Acesso em: 16 ago. 2015.

PICKLER, M. E. V. Web Semântica: ontologias como ferramentas de representação do conhecimento. *Perspect. ciênc. inf. [online]*, v. 12, n. 1, p. 65-83, 2007. Disponível em: <<http://www.scielo.br/pdf/pci/v12n1/05.pdf>>. Acesso em: 16 ago. 2015.

RAP. RDF API for PHP V0.9.6. 2014. Disponível em: <<http://wifo5-03.informatik.uni-mannheim.de/bizer/rdfapi/index.html>>. Acesso em: 20 nov. 2014.

RDF Schema 1.1. 2014. Disponível em: <<http://www.w3.org/TR/rdf-schema/>>. Acesso em: 6 dez. 2014.

REZENDE, S. O. *Sistemas inteligentes: fundamentos e aplicações*. Barueri, SP: Manole, 2005.

SANTOS NETO, Antonio Laurindo dos; et al. Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: um caso machadiano. *Transinformação*, v. 25, n. 1, p. 81-87, 2013.. Disponível em: <<http://www.scielo.br/pdf/tinf/v25n1/a08v25n1.pdf>>. Acesso em: 1º mar. 2015.

SCIELO. *Transinformação*. Disponível em: <<http://www.scielo.br/tinf>>. Acesso em: 1º mar. 2015.

SOUZA, Renato Rocha; ALVARENGA, Lídia. A Web Semântica e suas contribuições para a ciência da informação. *Ciência da Informação*, Brasília, v. 33, n. 1, p. 132-141, 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33n1/v33n1a16.pdf>>. Acesso em: 16 ago. 2015.