

# Anotações Semânticas em Repositórios Acadêmicos

Um estudo de caso com o RI UFBA

Aline Meira Rocha  
Universidade Federal da Bahia (UFBA)  
Salvador - BA - Brasil  
aline.meira@ufba.br

Lais do Nascimento Salvador  
Universidade Federal da Bahia (UFBA)  
Salvador - BA - Brasil  
laisns@ufba.br

Marlo V. dos Santos e Souza  
Universidade Federal da Bahia (UFBA)  
Salvador - BA - Brasil  
msouza1@ufba.br

## ABSTRACT

The current Web is syntactic and it is impossible for machines to process it. The Semantic Web with its technologies can bring more meaning to web content. UFBA's IR is on the Syntactic Web and the recovery of deposited items isn't always done efficiently. Information about the deposited items is stored in their metadata, but it is done manually and, not always, the chosen terms are able to help with the description process. The automatic semantic annotation method would certainly help librarians during the validation of the production's deposit as well as enabling the recovery of UFBA's IR items. The main proposal is to create a solution to semantically annotate the academic repository items in a semi-automatic method so to extract relevant terms that will be suggested to annotate the UFBA's IR items making it a semantic repository.

## KEYWORDS

semantic web, ontologies, natural language processing, institutional repositories, information extraction

## 1 INTRODUÇÃO

A Web Sintática não permite inferir conhecimento. Com isso a Web Semântica, uma extensão da Web atual, surge como alternativa para atribuir semântica às informações, permitindo que tanto pessoas quanto máquinas possam processá-las, trabalhando assim em cooperação [1]. Para dar significado aos dados, a Web Semântica faz uso de ontologias. As ontologias permitem que informações representadas por um mesmo conceito mas que utilizem diferentes identificadores sejam relacionadas, interligando esses termos entre si [1].

É possível extrair termos de documentos depositados em um repositório acadêmico, associá-los a uma ontologia e com isso gerar a anotação semântica desses documentos. Para isso, podem ser utilizadas diversas técnicas da Inteligência Artificial, incluindo o PLN (Processamento de Linguagem Natural - *Natural Language Processing*).

Nessa proposta são apresentados os resultados parciais e os próximos passos de uma proposta de dissertação, com o objetivo de desenvolver uma solução para anotar semanticamente de forma automatizada os itens de um RI (Repositório Institucional).

Esse artigo está dividido como segue: na seção 2 é abordada a fundamentação teórica e os trabalhos relacionados. A seção 3

trata da caracterização do problema. Enquanto a seção 4 aborda as contribuições esperadas. Na seção 5 pode ser vista a metodologia e o estado atual do trabalho. Já a seção 6 contém a descrição e avaliação dos resultados obtidos até o momento. Por fim, na seção 7 tem-se a conclusão e próximos passos.

## 2 FUNDAMENTAÇÃO TEÓRICA E TRABALHOS RELACIONADOS

### 2.1 Web Semântica

A Web Semântica, uma extensão da Web atual, permite adicionar significado ao conteúdo disponível, possibilitando que tanto pessoas quanto máquinas possam interpretá-los [1].

O termo ontologia surgiu na Filosofia [5]. Segundo [4], uma ontologia é definida como "uma explícita especificação de uma conceitualização". Já [5], afirma que o que existe é aquilo que pode ser representado, onde a ontologia é um mundo a ser modelado.

Para representar as ontologias é utilizada a linguagem OWL (*Web Ontology Language*). Existem 3 (três) variações da OWL: a OWL Full, a OWL DL (*Description Logic*) a OWL Lite, cada uma com um grau de expressividade. Atualmente, a OWL está em sua versão OWL 2 [19], com o objetivo de tornar a Web mais acessível às máquinas. Em relação à versão anterior foram adicionadas novas funcionalidades, como chaves, tipo de dados mais ricos e capacidade de anotações melhoradas.

A Web Semântica, através da Web de Dados, estabelece links entre dados de diversas fontes. A Web de Dados trata tudo como um recurso (pessoas, coisas, páginas web), que são identificados através de uma URI (*Uniform Resource Identifier*). Com isso, conteúdos que possuam relação, são ligados por tuplas RDF (*Resource Description Framework*) formando um imenso banco de dados global [2].

Anotações Semânticas possibilitam que os metadados de um documento sejam enriquecidos semanticamente, facilitando a sua recuperação. Assim, termos extraídos de um documento quando associados a uma ontologia (ou a conceitos dela) permitem descrever melhor esse documento [6].

As Anotações Semânticas podem ser de 4 (quatro) tipos: (1) manual (quando a anotação é realizada manualmente); (2) semi-automática (os termos são extraídos de forma automática e depois são sugeridos a um especialista para serem anotados); (3) automática (os termos são extraídos e anotados de forma automática) e (4) híbrida (quando ocorre mais de um tipo de anotação ao mesmo tempo). Essas Anotações podem ser tanto intrusivas (quando é feita no próprio documento) quanto não-intrusivas (quando é feita num arquivo à parte).

## 2.2 Repositórios Institucionais

Por sua vez, um RI serve para armazenar as produções científicas e acadêmicas geradas por uma instituição, possibilitando que seus pesquisadores possam divulgar as pesquisas realizadas [3] e aumentar sua visibilidade [12]. É necessário que seja feita uma divulgação dos resultados dessas pesquisas, como retorno aos investimentos feitos em ciência [16].

Atualmente, existem diversos *softwares* utilizados como solução de armazenamento para RI, sendo a mais utilizado o *DSpace*<sup>1</sup>, desenvolvido em uma parceria entre o MIT (*Massachusetts Institute of Technology*) e a HP (*Hewlett-Packard*) [18].

O *DSpace* se organiza em comunidades e sub-comunidades, seguindo a estrutura organizacional das instituições. Cada comunidade pode possuir inúmeras coleções que são compostas por itens e permitem armazenar arquivos em diversos formatos. O *DSpace* utiliza o padrão de metadados (dados sobre os dados) *Dublin Core*<sup>2</sup> (DC) e uma interface OAI (*Open Archives Initiative*) para colher os metadados armazenados [17].

Com o *DSpace* é possível realizar o auto-arquivamento, assim o próprio pesquisador pode realizar o depósito de sua publicação enquanto um bibliotecário fica responsável pela validação dos metadados informados [15]. Só depois dessa validação é que cada publicação fica disponível para os usuários finais.

## 2.3 Trabalhos Relacionados

Diversos trabalhos já foram publicados na área de Web Semântica com enfoque em anotação semântica, seja utilizando técnicas de PLN, seja sendo aplicada no âmbito dos repositórios acadêmicos.

O trabalho [13] apresenta uma proposta de uma técnica automática de anotação semântica para conteúdos multimídia. As limitações das técnicas automáticas apresentam uma alta dependência de inferências de bases de conhecimento e as abordagens manuais consomem tempo e apresentam dados incompletos. Diversos serviços possibilitam que os usuários possam adicionar *tags* para anotar o conteúdo produzido. Com isso, foi proposto um anotador onde é recebido um conjunto de *tags* que são filtradas gerando uma folksonomia (indexação de informação) de *tags*. Essas *tags* foram enriquecidas para posterior anotação seguindo 3 estratégias apresentadas no artigo. Por fim, o conteúdo foi anotado com os conceitos obtidos dessas estratégias.

O trabalho [7] descreve como um repositório no *Dspace* pode ser anotado semanticamente como um repositório semântico. Baseado no modelo de dados do *Dspace*, foi feito um mapeamento dos metadados dos conteúdos desse repositório (mapeamento R2RML), gerando um grafo RDF.

Já o artigo [14] aborda a extração automática de informações na área médica da hemoterapia utilizando ferramenta de PLN. Adotou-se para a extração de termos candidatos a ferramenta *Sketch Engine*, a construção de um corpus no domínio de sangue com a anotação *POS (Part-of-Speech) Tagger* (onde cada termo é etiquetado de acordo com seu uso na frase) e o cálculo das frequências dos termos candidatos à ontologia. Considerou-se somente os prefixos dos termos pois normalmente estes representam sua semântica. Após esse cálculo, os termos foram agrupados de acordo com a classe semântica

**Table 1: Trabalhos Relacionados**

Trabalho	Proposta	Limitações
[13]	Anotação semântica automática de conteúdo multimídia	Baseada em Folksonomia
[7]	Anotação semântica em repositório <i>Dspace</i>	Considera somente conteúdo armazenado em banco de dados
[14]	Extração automática de termos candidatos às ontologias na área médica	Processo sem validação de um especialista
[11]	Sistema de navegação no <i>Linked Data</i> numa ontologia <i>Dspace</i>	Considera somente conteúdo armazenado em banco de dados
[8]	Implementação de busca semântica no <i>Dspace</i>	Considera somente conteúdo armazenado em banco de dados

correspondente. Em seguida, foi gerada a taxonomia dessas classes semânticas e a validação da ontologia por um especialista na área.

No trabalho [11] é descrito um sistema de navegação para recursos educacionais no *Linked Data* mapeando esses recursos numa ontologia *Dspace*. Em [8] é proposto um modelo de inferência de base de conhecimento, recuperação e navegação. Para isso, a ontologia *Dspace* foi povoada e foi implementada uma busca semântica.

A tabela 1 apresenta uma síntese dos trabalhos relacionados, ressaltando suas propostas e limitações apresentadas.

Esse trabalho diferencia-se por propor o desenvolvimento de uma solução que permitirá a anotação semântica semi-automática. Termos serão sugeridos para serem anotados pelos especialistas, levando-se em consideração tanto os metadados armazenados em banco de dados quanto o conteúdo extraído de arquivos armazenados no repositório.

## 3 CARACTERIZAÇÃO DO PROBLEMA

Repositórios acadêmicos possibilitam o armazenamento das produções científicas de universidades e centros de pesquisa. Eles permitem o auto-arquivamento mas por mais que os metadados de cada publicação sejam preenchidos pelo próprio pesquisador, nem sempre os termos escolhidos para descrevê-la são os mais relevantes e adequados. Além disso, a falta de conhecimento multidisciplinar por parte dos bibliotecários, impacta na validação precisa desses metadados.

Extrair termos relevantes desses conteúdos, podem ajudar a descrevê-los melhor. O ideal seria que esses termos fossem sugeridos durante a validação dos metadados por parte dos bibliotecários.

## 4 CONTRIBUIÇÕES ESPERADAS

Nessa proposta espera-se desenvolver uma solução para anotar semanticamente de forma automatizada os itens de um RI. A anotação

<sup>1</sup> <http://www.dspace.org>

<sup>2</sup>DC ou DCMES (*Dublin Core Element Set*) composto por 15 (quinze) elementos

semântica automatizada auxiliaria o trabalho dos bibliotecários durante a validação dos depósitos das produções e ajudaria na recuperação das informações, otimizando a busca de itens pesquisados num RI.

Com isso pretende-se anotar semanticamente os itens de um RI no *Dspace* de maneira semi-automática. Para isso, serão sugeridos aos bibliotecários termos relevantes extraídos de textos em formato pdf (*Portable Document Format*) e de seus metadados, que servirão para anotarem semanticamente os itens de um repositório.

Baseando-se num repositório de informação que contém os itens a serem anotados, inicia-se o processo, conforme mostra a Figura 1 a seguir, realizando a extração de texto através de técnicas de Processamento de Linguagem Natural, onde será realizada a extração de *tokens* relevantes.

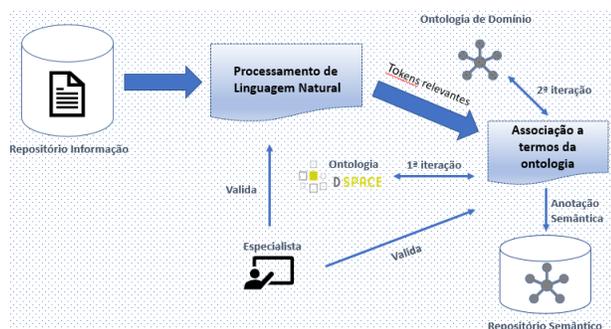


Figure 1: Esquema Anotação Semântica de Documentos. Fonte: elaborado pela autora.

Numa 1ª iteração, esses *tokens* serão associados a termos das ontologias *DSPACE* e *Dublin Core* que servirá para anotar os itens de acordo com a estrutura do repositório, ao passo que isso permitirá classificar melhor os itens depositados na comunidade/coleção correta, levando-se em consideração que um pesquisador pode ter mais de um vínculo com a instituição e acontecer dele informar erroneamente a comunidade ou coleção onde o item deve ser depositado. Depois disso, cada item depositado passa pela validação de um especialista, nesse caso um bibliotecário.

Já numa 2ª iteração, os *tokens* seriam associados a uma ontologia de domínio específico, já que o propósito dessa iteração é anotar os itens em relação ao seu conteúdo. Um especialista validará os termos sugeridos (processo iterativo), gerando por fim um repositório semântico.

Para atingir o objetivo geral dessa proposta de dissertação, foram definidos os seguintes objetivos específicos: (1) classificar os documentos nas comunidades/coleções a partir de termos relevantes extraídos; (2) avaliar os métodos de anotação semântica; (3) realizar a anotação semântica de forma automatizada dos documentos do RI; (4) anotar a estrutura dos itens de um RI utilizando as ontologias *Dspace* e *Dublin Core* (5) realizar um estudo de caso com o RI UFBA e (6) realizar a desambiguação com uma ontologia de domínio (a ser escolhida) e as ontologias *DSPACE* e *Dublin Core*.

## 5 ESTADO ATUAL DO TRABALHO

### 5.1 Metodologia

Para atingir os objetivos propostos nesse trabalho, foram definidos os seguintes passos metodológicos: (1) realização de levantamento bibliográfico; (2) estudo da estrutura de um repositório acadêmico; (3) pesquisa das ontologias *Dspace* e *Dublin Core*; (4) reuniões com equipes envolvidas com a gestão do RI UFBA; (5) montagem da base de dados para realização dos experimentos; (6) povoamento das ontologias encontradas; (7) avaliação dos classificadores; (8) avaliação dos métodos de anotação semântica; (9) geração do repositório semântico; (10) realização de um experimento com o pessoal do SIBI/UFBA; (11) análise dos resultados obtidos e (12) registro dos resultados obtidos.

### 5.2 Atividades em Andamento

Considerando-se os passos metodológicos informados na seção anterior, até o momento já foram realizados os seguintes passos (1), (2), (3) e (4). Ressalta-se que os passos (1) e (4) correspondem a atividades iterativas a serem realizadas no decorrer desse trabalho.

O passo (5) está em andamento e consiste na montagem do ambiente com o *dump* de uma visão do banco de dados em *Postgres* contendo as informações de produções depositados no RI UFBA e o *download* de 100 produções em formato pdf.

Como experimento inicial, realizou-se a extração de textos em um conjunto de arquivos pdf, escolhidos aleatoriamente do RI UFBA, com o auxílio da linguagem de programação *Python* e uma de suas bibliotecas para extração de texto, a *pdfminer*. Essa biblioteca permite extrair textos de arquivos em formato pdf e os converte para txt (arquivo de texto). Nos arquivos extraídos, notou-se que as imagens foram desconsideradas e que os textos apresentados em tabelas foram conservados, porém perdeu-se sua formatação. Esse experimento será replicado posteriormente num conjunto maior de arquivos.

Para realizar o processamento desses textos, será utilizada a biblioteca NLTK<sup>3</sup> (*Natural Language Toolkit*) codificada em *Python*. Essa biblioteca possui comandos para PLN executar as tarefas de tokenização (quebrar o texto em *tokens*, ou seja, em palavras) de sentenças, extração de prefixos ou sufixos, etiquetagem (definição da classe gramatical de uma palavra), identificação de entidades nomeadas (como pessoas, empresas ou instituições) remoção de *stopwords* (palavras pouco relevantes no texto) e outras tarefas.

O próximo passo será normalizar os textos obtidos, corrigindo possíveis inconsistências, como exclusão de linhas em branco, corrigir problemas de acentuação, tratar linhas com poucos caracteres, dentre outras correções.

## 6 DESCRIÇÃO E AVALIAÇÃO DOS RESULTADOS

Depois de identificar o problema e definir os objetivos desse trabalho, buscou-se mapear os passos metodológicos a serem seguidos. Com isso, para seguir o cronograma estipulado, algumas atividades já se encontram concluídas.

<sup>3</sup><http://nltk.org>

## 6.1 Realização de levantamento bibliográfico

Realizou-se uma pesquisa em diversas bases de dados por trabalhos desenvolvidos na área de pesquisa. A princípio, foi feita uma pesquisa por trabalhos na área de Ontologias e Anotações Semânticas que tenham como objeto de estudo, os Repositórios Institucionais (ou Acadêmicos). Também foram pesquisados trabalhos com a extração de texto utilizando técnicas de PLN.

Esses trabalhos servirão de apoio para a fundamentação teórica desse projeto, bem como para verificar o que está sendo pesquisado e desenvolvido nessas áreas, além de identificar as lacunas de pesquisa.

## 6.2 Estudo da estrutura de um repositório acadêmico

Estudou-se a estrutura de um repositório institucional acadêmico, como ele é organizado, mapeando suas principais características e funcionalidades. Verificou-se como foi feita a instalação do RI UFBA, a fim de verificar se foi feita alguma alteração na instalação padrão, como customizações, adição de plugins, etc. Foram verificadas quais as configurações do ambiente onde o RI UFBA se encontra instalado atualmente.

**6.2.1 A instalação do RI UFBA.** A versão do *DSpace* instalada atualmente no RI UFBA é a 3.2, hospedada num servidor *Linux Debian*, com banco de dados *PostgreSQL 8.3*, *Apache 2.2.16*, *Tomcat 7.0.47*, Java versão *JDK (Java SE Development Kit) 1.5*, com pacotes *Java-Ant* e *Maven 2* instalados, utilizando as estatísticas da Universidade do Minho.

A UFBA (Universidade Federal da Bahia) armazena uma parte de sua produção científica e acadêmica em seu repositório institucional, o RI UFBA. O RI UFBA foi implantado pela EDUFBA (Editora da UFBA) em 2010 e pode ser acessado no endereço <https://repositorio.ufba.br> [16], sendo o *DSpace* instalado num servidor exclusivo fornecido pelo IBICT (Instituto Brasileiro de Informação em Ciência e Tecnologia).

Suas comunidades e sub-comunidades estão organizadas de acordo com a estrutura organizacional da UFBA e os depósitos são realizados através do auto-arquivamento, tendo os metadados validados pelos bibliotecários ligados à comunidade onde a publicação foi depositada.

## 6.3 Pesquisa das ontologias *DSpace* e *Dublin Core*

Realizou-se uma pesquisa com a finalidade de verificar se já existiam ontologias publicadas sobre o *DSpace* e o padrão de metadados *Dublin Core*. Essa busca incluiu desde repositórios de ontologias a trabalhos publicados sobre o tema.

As ontologias encontradas<sup>4</sup> foram criadas pelo grupo de pesquisa *Semantic Web and Interoperability* da Universidade de Patras [9] [10]. A Figura 2 ilustra a taxonomia da ontologia *DSpace* encontrada:

Conforme ilustrado na Figura 2, as seguintes classes foram identificadas:

- *Collection*: representando as coleções

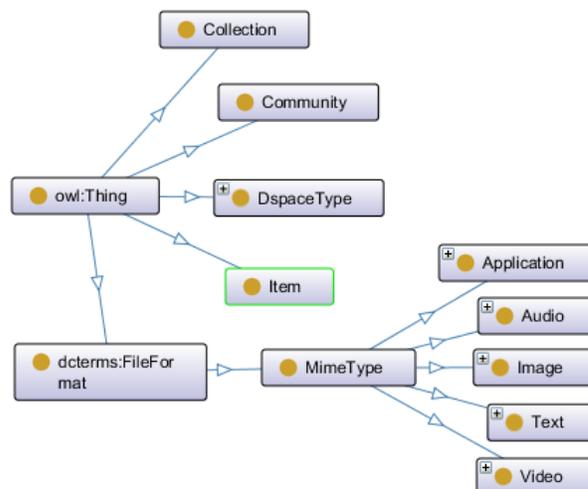


Figure 2: Ontologia *DSpace* (extraída do Protégé)

- *Community*: representando as comunidades, refletindo a divisão hierárquica existente dentro das instituições.
- *DspaceType*: representando os tipos de documentos que podem ser depositados
- *dcterms:FileFormat* (com a sub-classe *MimeType*): representando os formatos dos arquivos suportados. A sub-classe *MimeType* possui uma partição de 5 (cinco) sub-classes: *Application*, *Audio*, *Image*, *Text* e *Video*, referindo-se aos tipos de arquivos que podem ser depositados
- *Item*: representando as publicações depositadas

As propriedades existentes entre essas classes são descritas com maiores detalhes na Tabela 2 abaixo:

Table 2: Classes e Propriedades - Ontologia *DSpace*.

Classes	Propriedades
Collection	(not (Item)) and (dcterms:hasPart some Item) SubClassOf Collection / disjoint with item
Community	disjoint with item
MimeType	SubclassOf of dcTerms:FileFormat has subclasses Application, Audio, Image, Text and Video
Item	(not (Item)) and (dcterms:hasPart some Item) SubClassOf Collection / disjoint with Community and Collection

Por exemplo, a propriedade *(not (Item)) and (dcterms:hasPart some Item) SubClassOf Collection* - presente nas classes *Item* e *Collection* - define que um item é parte de uma coleção menos item dele mesmo e que coleções só podem conter itens.

Enquanto a ontologia *dspace-ont.owl* possui as classes, representando a estrutura de um repositório institucional, a ontologia *dc-ont.owl* apresenta as propriedades que ligam essas classes, conforme mostra a Tabela 3 a seguir:

Assim, um *Item* faz parte de uma *Collection* (*dcterms:isPartOf*) e o inverso, onde uma *Collection* contém um determinado *Item* (*dcterms:hasPart*).

<sup>4</sup> <https://github.com/swigroup/dspace-semantic-search.ontologies>

Table 3: Propriedades DCterms.

Propriedade	Inversa
dcterms:hasFormat	dcterms:isFormatOf
dcterms:hasPart	dcterms:isPartOf
dcterms:hasVersion	dcterms:isVersionOf
dcterms:isReferencedbBy	dcterms:references
dcterms:isReplaceBy	dcterms:replaces
dcterms:isRequiredBy	dcterms:requires

Já a Tabela 4, mostra a relação entre as propriedades simétricas da ontologia *Dublin Core*:

Table 4: Relação Simétrica.

Propriedade	Simétrica
dcterms: relation	dcterms: relation

Nesse caso, a propriedade *dcterms: relation* é simétrica a si mesma com a finalidade de descobrir relações entre indivíduos.

A ligação entre as duas ontologias é feita através dos *dcterms*<sup>5</sup> ou QDC (*Qualified Dublin Core*), qualificadores que representam as relações existentes e servem para refinar os elementos *Dublin Core*.

#### 6.4 Reuniões com equipes envolvidas com a gestão do RI UFBA

Periodicamente serão realizadas reuniões com as equipes responsáveis pelo RI UFBA. A gestão do RI UFBA atualmente é realizada pelo Núcleo Científico-SIBI/UFBA (Sistema Universitário de Bibliotecas) e a CRI-STI/UFBA (Coordenação de Redes e Infraestrutura - Superintendência de Tecnologia da Informação) é responsável pela infraestrutura do ambiente que hospeda o RI UFBA.

#### 6.5 Montagem da base de dados para realização dos experimentos

A base de dados para realização dos experimentos será composta do *dump* de uma visão do banco de dados em *PostgreSQL* onde estão armazenadas as informações das publicações depositadas no RI UFBA e os *downloads* de 100 publicações em formato pdf.

Será realizado um recorte (*dataset*) desse *dump*, selecionando um conjunto de dados armazenado no RI e depois serão realizados diversos experimentos com esse recorte.

Serão desconsiderados registros relacionados à configuração da ferramenta e às informações de cadastro de usuários.

## 7 CONCLUSÕES E PRÓXIMOS PASSOS

Esse trabalho descreveu uma proposta que tem como objetivo desenvolver uma solução para anotar semanticamente de forma automatizada os itens de um RI, realizando um Estudo de Caso com o RI UFBA. Para isso foram descritos os objetivos e a metodologia a ser utilizada no trabalho. Além disso, foram descritos os resultados parciais desse projeto.

<sup>5</sup><http://dublincore.org/documents/usageguide/qualifiers.shtml>

A anotação semântica de itens de um RI pode favorecer a recuperação das informações, otimizando a busca de itens pesquisados num RI. Essa anotação sendo feita de maneira semi-automática poderá auxiliar o trabalho dos bibliotecários na validação dos metadados de cada item depositado no RI.

Como próximos passos, será finalizada a montagem do ambiente para realização dos experimentos. Em seguida, a extração dos termos com o PLN será retomado a fim de extrair os *tokens* relevantes dos arquivos. Também será realizada uma avaliação dos classificadores, dos métodos de anotação semântica e realização de um experimento com a anotação dos itens pela equipe de bibliotecários. Por fim, será feita uma análise e registro dos resultados obtidos com a geração de um repositório semântico.

## REFERENCES

- [1] Tim Berners-Lee, James Hendler, and Ora Lassila. 2001. **The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities.** *Scientific American* 284, 5 (2001), 1–5. [http://csis.pace.edu/~marchese/CS835/Lec9/112\\_SemWeb.pdf](http://csis.pace.edu/~marchese/CS835/Lec9/112_SemWeb.pdf)
- [2] Christian Bizer, Tom Heath, and Tim Berners-Lee. 2009. **Linked data: the story so far.** *Semantic services, interoperability and web applications: emerging concepts* (2009), 205–227. <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>
- [3] M. J. Gomes and F. Garcia Rosa. 2010. **Repositórios Institucionais.** EDUFBA, Brasil. <https://repositorio.ufba.br/ri/bitstream/ri/616/3/Repositorios%20institucionais.pdf>
- [4] Thomas R Gruber et al. 1993. **A translation approach to portable ontology specifications.** *Knowledge acquisition* 5, 2 (1993), 199–220. <http://tomgruber.org/writing/ontologia-kaj-1993.pdf>
- [5] Nicola Guarino, Daniel Oberle, and Steffen Staab. 2009. **What is an Ontology?** In *Handbook on ontologies*. Springer, 1–17. [http://iaoa.org/isc2012/docs/Guarino2009\\_What\\_is\\_an\\_Ontology.pdf](http://iaoa.org/isc2012/docs/Guarino2009_What_is_an_Ontology.pdf)
- [6] Atanas Kiryakov, Borislav Popov, Ivan Terziev, Dimitar Manov, and Damyan Ognyanoff. 2004. **Semantic annotation, indexing, and retrieval.** *Web Semantics: Science, Services and Agents on the World Wide Web* 2, 1 (2004), 49–79.
- [7] Nikolaos Konstantinou, Dimitrios-Emmanuel Spanos, Nikos Houssos, and Nikolaos Mitrou. 2014. **Exposing scholarly information as linked open data: RDFizing DSpace contents.** *The Electronic Library* 32, 6 (2014), 834–851.
- [8] Dimitrios A Koutsomitropoulos, Georgia D Solomou, Andreas D Alexopoulos, and Theodore S Papatheodorou. 2009. **Knowledge Management and Acquisition in Digital Repositories - A Semantic Web Perspective.** In *KMIS*. Citeseer, 117–122. <http://www.hpclab.ceid.upatras.gr/viografika/kotsomit/pubs/kmis09.pdf>
- [9] Dimitrios A. Koutsomitropoulos, Georgia D. Solomou, Andreas D. Alexopoulos, and Papatheodorou Theodore S. 2009. **Semantic Metadata Interoperability and Inference-Based Querying in Digital Repositories.** (2009). <http://www.hpclab.ceid.upatras.gr/viografika/kotsomit/pubs/jitr09.pdf>
- [10] Dimitrios A. Koutsomitropoulos, Georgia D. Solomou, Andreas D. Alexopoulos, and Papatheodorou Theodore S. 2010. **Semantic Web enabled digital repositories.** (2010). [http://web.simmons.edu/~weigle/INTERO\\_METADATA\\_LANGUAGE/Koutsomitropoulos\\_SemanticWeb.pdf](http://web.simmons.edu/~weigle/INTERO_METADATA_LANGUAGE/Koutsomitropoulos_SemanticWeb.pdf)
- [11] Dimitrios A Koutsomitropoulos, Georgia D Solomou, and Aikaterini K Kalou. 2014. **Navigating among Educational Resources in the Web of Linked Data.** In *International Conference on Web Reasoning and Rule Systems*. Springer, 185–190. [http://rd.springer.com/chapter/10.1007/978-3-319-11113-1\\_14](http://rd.springer.com/chapter/10.1007/978-3-319-11113-1_14)
- [12] Fernando Leite, Bianca Amaro, Tainá Batista, and Michelli Costa. 2012. **RI - Repositórios Institucionais.** IBICT, Brasil. <http://livroaberto.ibict.br/handle/1/703>
- [13] Marcelo Garcia Manzato and Rudinei Goularte. 2012. **Automatic annotation of tagged content using predefined semantic concepts.** In *Proceedings of the 18th Brazilian symposium on Multimedia and the web*. ACM, 237–244.
- [14] Fabrício Mendonça, Mauricio Barcellos Almeida, Renato Rocha Souza, and Daniela L Silva. 2012. **Extração Automática de Termos Candidatos às Ontologias: um Estudo de Caso no Domínio da Hemoterapia.** *ONTOBRAS-MOST* 938 (2012), 170–175.
- [15] Flavia Rosa, Rodrigo França Meirelles, and Marcos Palacios. 2011. **Repositório Institucional da Universidade Federal da Bahia: implantação e acompanhamento.** (2011). <https://repositorio.ufba.br/ri/bitstream/ri/1590/1/5603.pdf>
- [16] Luis Sayão, Lídia Brandão Toutain, Flavia Garcia Rosa, and Carlos Henrique Marcondes. 2009. **Implantação e gestão de repositórios institucionais.** EDUFBA, Brasil. [https://repositorio.ufba.br/ri/bitstream/ufba/473/3/implantacao\\_repositorio\\_web.pdf](https://repositorio.ufba.br/ri/bitstream/ufba/473/3/implantacao_repositorio_web.pdf)

- [17] Milton Shintaku and Rodrigo França Meirelles. 2010. **Manual do DSpace: administração de repositórios**. (2010). [https://repositorio.ufba.br/ri/bitstream/ri/769/1/Manual%20do%20Dspace\(2\).pdf](https://repositorio.ufba.br/ri/bitstream/ri/769/1/Manual%20do%20Dspace(2).pdf)
- [18] Robert Tansley, Mick Bass, David Stuve, Margret Branschofsky, Daniel Chudnov, Greg McClellan, and MacKenzie Smith. 2003. **The DSpace institutional digital repository system: current functionality**. In *Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries*. IEEE Computer Society, 87–97.
- [19] W3C. 2012. **OWL 2 Web Ontology Language**. (2012). <https://www.w3.org/TR/owl2-overview/>