

ANOTAÇÃO SEMÂNTICA DE CONTEÚDO NO APOIO A PRODUÇÃO DE OBJETOS DE APRENDIZAGEM

Ronaldo Amaral Santos – rsantos@iff.edu.br
Instituto Federal de Educação, Ciência e Tecnologia Fluminense
Rua Doutor Siqueira, 273 Parque Dom Bosco
28030-130 – Campos dos Goytacazes - RJ

Georgia R. R. Gomes – georgia@ucam-campos.br
Universidade Candido Mendes
Rua Anita Peçanha, 100 Parque São Caetano
28030-335 – Campos dos Goytacazes - RJ

Mark Douglas de A. Jacyntho – markjacyntho@gmail.com
Universidade Candido Mendes
Rua Anita Peçanha, 100 Parque São Caetano
28030-335 – Campos dos Goytacazes - RJ

Resumo: *Na Educação à distância – EAD, um dos elementos vitais é a disponibilização de materiais de aprendizagem relevantes e de valor. Nos últimos anos, muitas aplicações educacionais baseadas na Web têm sido desenvolvidas, mas alguns desafios ainda existem, dentre os quais a pesquisa por materiais e objetos de aprendizagem mais inteligentes e eficientes. Para tal, pode-se utilizar-se dos conceitos introduzidos com a Web Semântica, onde metadados estruturados inteligíveis por máquina são adicionados à Web de forma a agregar significado à informação. A agregação de valor semântico aos documentos é uma forma de organizar o processo de publicação, recuperação e enriquecimento da informação, haja vista que possibilita a recuperação da informação contida nos documentos de forma precisa e eficaz, facilitando sua reutilização. Portanto, neste artigo é apresentado um protótipo baseado em uma arquitetura de serviços web para anotação semântica de conteúdo, que automatiza a extração de conceitos em objetos de aprendizagem, utilizando técnicas de mineração de texto, com estudo de caso aplicado ao ambiente de aprendizagem Moodle. Com este trabalho espera-se construir novas formas de descoberta de conhecimento e reuso de informação, a partir de dados disponíveis em ambientes de aprendizagem, por meio da descrição estruturada em RDF dos objetos de aprendizagem. Além disso, interligar conceitos a fonte de dados na Web de Dados (Web of Linked Data), enriquecendo a base de conhecimento do ambiente de aprendizagem com mashup semântico entre estes conceitos e recursos pré-existentes da Web de dados, possibilitando, um aumento da produtividade no processo de ensino-aprendizagem.*

Palavras-chave: *Anotação Semântica, EAD, Objeto de Aprendizagem, Linked Data*



1 INTRODUÇÃO

A partir do crescimento da modalidade de ensino à distância – EAD em conjunto com a evolução das Tecnologias da Informação (TI) tornou-se importante incentivar essa modalidade de ensino por meio da pesquisa por novas tecnologias para suporte ao ensino EAD. A Internet impulsiona esse movimento, pois facilita o compartilhamento de documentos eletrônicos, independente de sistema operacional, hardware ou dispositivo (GOMES, 2006).

Um dos elementos vitais no ensino à distância é a disponibilização de materiais de aprendizagem relevantes e de valor. É sempre recomendável que seja valorizado o reaproveitamento e a interoperabilidade entre diferentes plataformas na produção de conteúdos e recursos didáticos de qualidade, pois estes itens representam um custo elevado no processo de ensino e aprendizagem. (MOURA, 2005).

Segundo Araujo (2003), a Web está se tornando uma grande biblioteca virtual, onde a informação sobre qualquer assunto está disponível a qualquer hora e em qualquer lugar, com ou sem custo, criando oportunidades em várias áreas do conhecimento humano, dentre as quais a Educação. Porém, as informações na Web não são estruturadas e organizadas, as máquinas não podem “compreender” e nem “interpretar” o significado das informações. Embora muitas aplicações educacionais baseadas na Web tenham sido desenvolvidas nos últimos anos, alguns problemas nesta área não foram bem resolvidos, entre os quais está a pesquisa de materiais e objetos de aprendizagem de forma mais inteligente e precisa.

Em Berners-Lee et al. (2001) foi proposta uma extensão da Web convencional, onde metadados estruturados inteligíveis por máquina são adicionados à Web, de forma que computadores possam entender o significado da informação publicada e, portanto, executar automaticamente, em larga escala, as tarefas que são executadas manualmente. Esta nova Web que agrega significado a informação é denominada Web Semântica.

A agregação de valor semântico aos documentos, proposta pela Web semântica, é uma forma de organizar o processo de publicação, recuperação e enriquecimento da informação. As anotações semânticas, ou seja, as associações das expressões relevantes de trechos de textos ou metadados descrevendo os documentos a conceitos e instâncias descritos em um domínio podem permitir que a recuperação da informação contida nos documentos seja realizada com maior precisão e eficácia, facilitando a sua reutilização. Muito embora, outros problemas precisam ser solucionados para alcançar este objetivo, entre eles estão a definição de métodos e ferramentas para automatizar o processo de anotação semântica.

Boa parte da automatização do processo de anotação semântica está relacionada com a descoberta de conceitos em conteúdos textuais. Para que esta descoberta possa ser automatizada são utilizadas técnicas de Processamento de Linguagem Natural (PLN). PLN é uma área de pesquisa cujo objetivo é estudar o desenvolvimento de técnicas e ferramentas que analisam, reconhecem ou geram textos em linguagens humanas ou linguagens naturais. Para Lopes (2011), PLN é uma área com grandes desafios pois a linguagem natural é rica em ambigüidades, diferentemente das linguagens formais que são definidas evitando a ambigüidade.

Com base nos conceitos apresentados, o objetivo geral deste trabalho é propor um sistema que automatize a extração de conceitos em objetos de aprendizagem, baseando-se em técnicas de mineração de texto e realizando a anotação semântica automatizada destes conteúdos, a partir de um ambiente de aprendizagem.

Estas anotações são utilizadas para a descrição explícita dos objetos de aprendizagem, por meio de metadados estruturados em *Resource Description Framework* (RDF), instanciando ontologias de referência e em conformidade com os princípios *Linked Data*, enunciados em Berners-Lee (2006). O objetivo é permitir que a máquina seja capaz de recuperar a informação de forma mais precisa, com base nas relações ontológicas explícitas entre os conceitos presentes nos conteúdos.

A fim de enriquecer sobremaneira a base de conhecimento do ambiente de aprendizagem (e, por conseguinte, aumentar o poder e a abrangência das buscas), os relacionamentos entre os conceitos presentes nos objetos de aprendizagem, são enriquecidos por meio do *mashup* semântico entre estes conceitos e recursos pré-existentes da Web de dados (*Web of Linked Data*). Para o *mashup* semântico, dentre as fontes de dados presentes na Web de Dados, a escolhida foi a DBpedia¹ (Wikipedia em RDF), por dois motivos, a saber: ser uma fonte de dados central de referência e por descrever múltiplos domínios de conhecimento (*crossdomain*).

Para o estudo de caso foi desenvolvido um *web service* para anotação semântica de conteúdo e armazenamento dos relacionamentos semânticos em um repositório RDF. Os serviços são integrados ao ambiente de aprendizagem Moodle, por meio de *plugins* clientes do *web service*, que permitem a anotação semântica e buscas dos objetos de aprendizagem diretamente na plataforma, de forma transparente para o usuário.

A seguir serão apresentadas de forma breve algumas definições consideradas importantes para o entendimento do trabalho proposto.

2 FUNDAMENTAÇÃO

Nesta seção será abordada uma revisão de literatura indispensável ao entendimento e desenvolvimento do trabalho.

2.1 Mineração de Texto

Mineração de texto (*Text Mining* ou *Knowledge Discovery from Texts* - KDT) pode ser definida como o nome dado às técnicas de análise e extração de dados a partir de textos, frases ou apenas palavras (GOMES, 2006). Lopes (2004) afirma que Mineração de Textos pode também ser definida como um conjunto de técnicas e processos que se prestam a descobrir conhecimento inovador nos textos.

Para Gomes (2006) por meio da análise de textos é possível a descoberta de conceitos, classificações automatizadas e sumarizações para documentos não estruturados.

Em Monteiro et al. (2006) são descritas as três principais etapas do *Text Mining*: a etapa de Pré-Processamento ou preparação dos dados, a etapa de Análise dos Dados e Extração do Conhecimento, também conhecida como Etapa de Processamento de Textos e a etapa de Pós-processamento ou Avaliação das Descobertas.

A Figura 1 demonstra o processo de mineração de texto que se inicia na etapa de Pré-processamento, envolvendo a aplicação das técnicas de Processamento de Linguagem Natural PLN, *Stemming* e remoção de *Stopwords*. A segunda etapa é o Processamento de Textos, onde podem ser aplicados métodos de Extração de informação, Sumarização, Categorização e Clusterização. A última etapa, Pós-processamento ou Avaliação das Descobertas, apresenta os

¹ DBpedia - <http://dbpedia.org>

resultados obtidos na fase anterior, baseando-se nas regras definidas para a base de conhecimento.

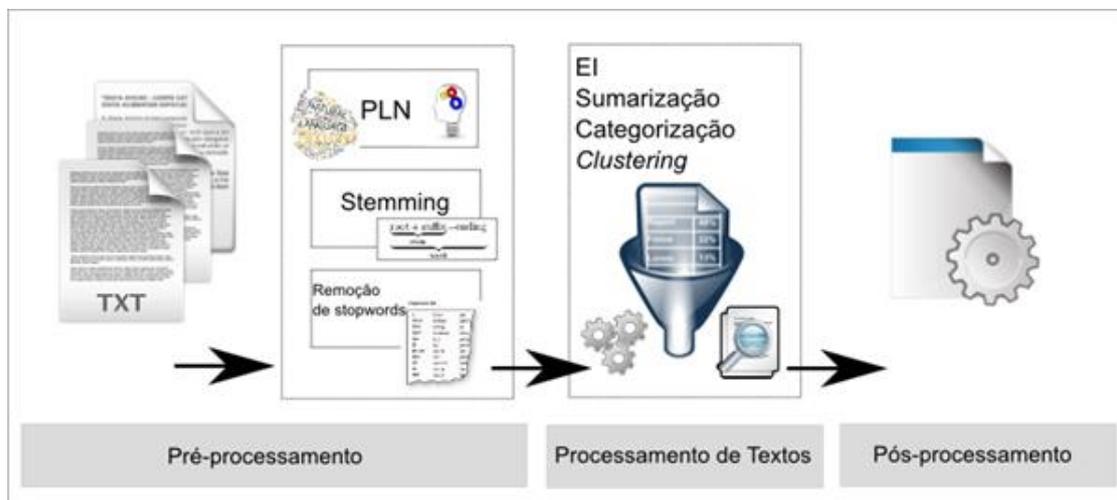


Figura 1 – Etapas do processo de Mineração de Texto.

2.2 Web Semântica

A Web Semântica não é uma Web separada, mas uma extensão da atual, na qual a informação é utilizada com significado bem definido, aumentando a capacidade dos computadores para trabalharem em cooperação com as pessoas (BERNERS-LEE et al., 2001).

Para Araujo (2003) é uma possibilidade de ter dados na Web conectados e com significados definidos, de modo a serem usados pelos computadores. Por exemplo, se em determinada página Web existir a palavra “banco” será possível distinguir se ela significa um “assento” ou um “estabelecimento comercial”. Neste contexto, pode-se perceber que a Web Semântica pode facilitar e melhorar a recuperação de informações relevantes, já que a própria máquina, dotada de ferramentas inteligentes, pode identificar o conteúdo de um *site*, por associação e dedução automática, antes de trazê-lo ao usuário como resultado de uma pesquisa (PICKLER, 2006).

Para que as informações possam ser compreendidas tanto por humanos quanto por computadores Berners-Lee et al. (2001) propõe os padrões da Web Semântica, representados na Figura 2, onde é definida uma arquitetura em camadas.

Esta arquitetura define a representação sintática, estrutural, semântica e lógica de informações referentes aos recursos Web. Para atingir os propósitos da Web Semântica, Souza & Alvarenga (2004) observam que é necessária uma padronização de tecnologias, de linguagens e de metadados descritivos, de forma que esta possa ser consumida por outros usuários, de uma maneira automática e não ambígua.

A primeira camada denominada URI (*Uniform Resource Identifier*) / IRI (*Internationalized Resource Identifiers*), definida na arquitetura da Web Semântica, permite que recursos (objetos de dados) disponíveis na Web sejam identificados de forma global, usando o mesmo esquema de endereços do protocolo HTTP, já consagrado na Web convencional. Por exemplo, poderíamos ter um URI que identifica o autor deste artigo

(<http://www.ucam-campos.br/pessoas/ronaldoamaral>), bem como um URI que identifica o próprio artigo (<http://www.abenge.org.br/cobenge-2014/artigos/1234>).

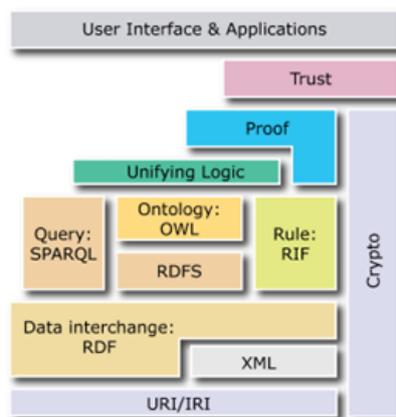


Figura 2 - Arquitetura em camadas da Web Semântica (HAWKE et al., 2013).

A camada XML (*Extensible Markup Language*) em conjunto com XML *namespace* e XML *schema*, garantem uma definição de sintaxe comum a ser usada na web semântica. XML é uma linguagem de marcação para documentos contendo informação estruturada.

A principal forma de representação na Web Semântica é o *Resource Description Framework* (RDF). O RDF é um *framework* para representação de informação acerca de recursos em forma de grafos, baseando-se em triplas "sujeito-predicado-objeto" (ou recurso-propriedade-valor). Por exemplo, para representar a informação que a pessoa identificada pelo URI "<http://www.ucam-campos.br/pessoas/ronaldoamaral>" é autor do artigo correspondente ao URI "<http://www.abenge.org.br/cobenge-2014/artigos/1234>", bastaria criar um link entre estes dois URIs e associar o rótulo "é autor de" ao link. Este rótulo, na verdade, também é um URI "<http://example.org/ontologia/autorOf>" que identifica uma propriedade definido em alguma ontologia.

2.3 Web de Dados – *Linked Data*

Como parte do movimento da Web Semântica, encontra-se em Berners-Lee (2006) a definição do conceito de *Linked Data* (em português Dados Ligados), que são um conjunto de diretrizes para publicar e conectar dados estruturados na Web, formando a chamada *Web of Linked Data* (Web de Dados), cuja topologia é ilustrada na Figura 3. Cada círculo representa um conjunto de dados em RDF (datasets) publicado segundo os princípios *Linked Data*. Os links entre os conjuntos de dados indicam a existência de pelo menos uma tripla com o sujeito no *dataset* de origem e com o objeto no *dataset* de destino, ou seja, *mashup* semântico entre fontes de dados. É importante destacar o conjunto de dados Dbpedia, apresentado no centro da Figura 3, por ser uma fonte de dados central de referência e por descrever múltiplos domínios de conhecimento (*crossdomain*).

A ideia é usar a arquitetura pré-existente da Web não apenas para publicar e interligar documentos, mas também para publicar e interligar (relacionar) dados. Em outras palavras, publicar dados diretamente na Web, identificando-os por URIs e estabelecer *links* (relacionamentos) entres estes dados por meio de URIs que representam propriedades descritas em ontologias. Tudo isso utilizando o modelo de dados RDF, formando um único

determinado domínio de conhecimento e que as anotações semânticas, que venham a descrever a relação de recursos na Web e as instâncias de uma ontologia, sejam escritas utilizando o modelo RDF (BECHHOFFER et al., 2004).

A Figura 4 de Popov et al. (2003), descreve a ideia de anotação semântica em conteúdo textual como a atribuição de *links* com a descrição semântica em entidades textuais. A ideia deste tipo de metadados é fornecer a ambos, classe e instância, informações sobre as entidades nos referidos documentos.

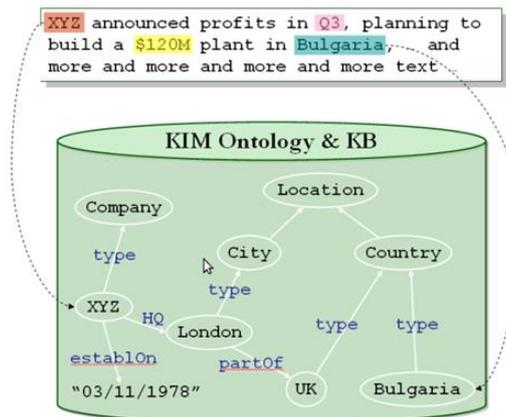


Figura 4 – Anotação Semântica (POPOV et al., 2003).

Oren (2006) destaca que existem várias ferramentas e paradigmas para a criação de anotações semânticas em recursos Web, sendo elas de forma manual, semi-automática ou totalmente automática.

3 PROTÓTIPO DESENVOLVIDO E ESTUDO DE CASO

Nesta seção será apresentado o protótipo desenvolvido baseado em uma Arquitetura Orientada a Serviços – SOA que visa realizar a extração de conceitos, anotação semântica e armazenamento em repositório RDF. Inicialmente, será feita a descrição do *web service* desenvolvido e em seguida a implementação do estudo de caso no ambiente de aprendizagem. O objetivo deste estudo de caso é aplicar a utilização dos serviços construídos ao ambiente de aprendizagem Moodle na versão 2.4 e à fonte de dados do DBpedia. Um dos benefícios desta arquitetura é possibilitar integrações futuras com outros ambientes de aprendizagem.

3.1 Protótipo do Web Service

O protótipo utiliza a forma de comunicação cliente-servidor implementado via Transferência de Estado Relacional (REST), que pode ser definido como um padrão arquitetural para projetos de *web services* onde o foco está no acesso aos recursos simples, identificados por URIs únicas, e sem estado utilizando-se os métodos (GET, POST, PUT e DELETE) do protocolo HTTP, onde cada recurso pode ter uma ou mais representações (XML, JSON, Text, etc) as quais são transferidas entre o cliente e o serviço, durante a invocação ao método (KAMALELDIN & DUMINDA, 2012).

A base do *web service* é composta pelos seguintes módulos:

- Servidor RESTful – responsável por prover as APIs necessárias e o controle aos acessos aos diversos módulos do serviço.
- Extração de Conceitos – responsável por extrair conceitos de conteúdo textual não estruturado utilizando técnicas de mineração de texto.
- Anotação Semântica – responsável pela transformação dos conceitos extraídos em anotações semânticas e a ligação destes conceitos com outras fontes de dados da Web de Dados.
- Armazenamento – responsável por persistir os dados e as anotações semânticas em repositório específico para armazenamento de triplas RDF – *RDF Triplestore*
- Consulta – responsável por disponibilizar uma interface amigável de recuperação ao conteúdo anotado, utilizando, sem que o usuário se dê conta, de forma encapsulada, a linguagem de consulta estruturada padrão da Web Semântica SPARQL³.

A Figura 5 apresenta uma visão geral da arquitetura do *web service* e sua integração com os componentes do sistema.

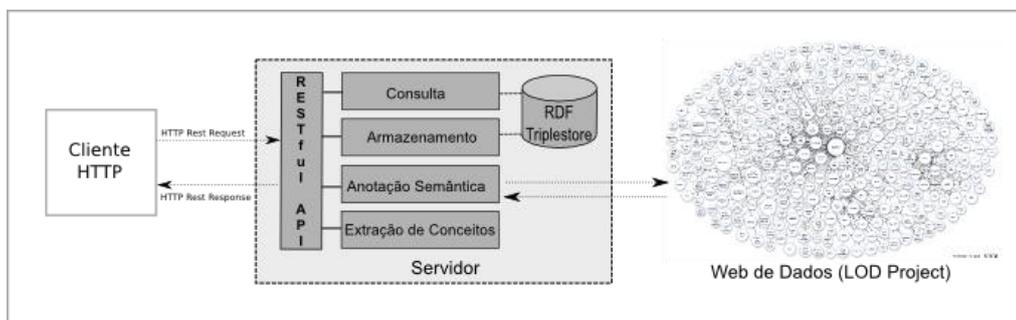


Figura 5 - Visão geral da arquitetura do *web service*.

3.2 Estudo de caso no ambiente Moodle

O ambiente de aprendizagem Moodle na versão 2.4 foi customizado através da instalação e desenvolvimento de *plugins* que possibilitam a comunicação com os diversos serviços providos pelo *web service*, possibilitando uma visão integradora dos serviços, conforme apresentado na Figura 6.

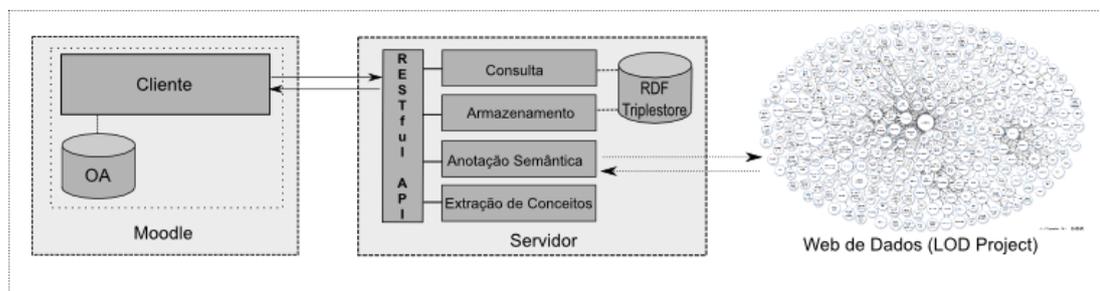


Figura 6 - Visão geral da integração Moodle e *web service*.

³ SPARQL - <http://www.w3.org/TR/sparql11-query/>

O fluxo de comunicação do sistema e a interação entre os módulos da arquitetura orientada a serviços são apresentados na Figura 7, onde é demonstrado o fluxo da anotação semântica de conteúdo distribuído em três camadas: Aplicação, Servidor e Web de dados.

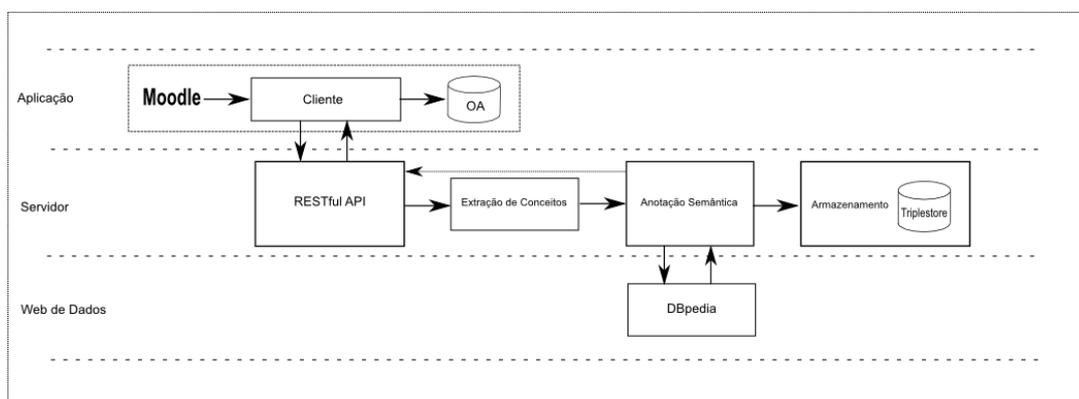


Figura 7 - Fluxo da interação entre as camadas da arquitetura SOA.

Na camada de Aplicação, encontra-se o ambiente de aprendizagem Moodle, que realiza a comunicação com a camada Servidor por meio de três *plugins* clientes específicos desenvolvidos para a *api restfull* disponibilizada pelo *front-end* do *web service*. São eles:

- **moodle-repository-semantic-lo**: Responsável pela interface de consulta e acesso aos objetos armazenados.
- **moodle-repository-semantic-lo-upload**: Responsável pela inserção de objetos de aprendizagem com conteúdo a ser armazenado no repositório local.
- **moodle-repository-semantic-lo-url**: Responsável pela inserção de objetos de aprendizagem com referência a recursos externos ao repositório.

Ao ser criado um objeto de aprendizagem no ambiente Moodle, este é processado pelo *plugin* cliente específico antes de seu armazenamento final, para que seu conteúdo seja anotado semanticamente. O usuário escolhe o tipo de recurso que será armazenado, referências ou arquivos, faz o preenchimento dos metadados específicos e envia à camada servidor os dados a serem processados.

A camada Servidor realiza o processamento deste objeto de aprendizagem criado e retorna o seu conteúdo anotado semanticamente.

Na primeira etapa é realizada a Extração de Conceitos, que identifica os conceitos relevantes no conteúdo, por meio de técnicas de mineração de textos e envia estas informações ao módulo de Anotação Semântica. Na Figura 8, é apresentada a tela de inserção do metadado de descrição com a identificação dos conceitos relevantes e a apresentação para os usuários dos conceitos encontrados no DBpedia, a fim de proporcionar a marcação semi-automática do conteúdo.

Na etapa de Anotação Semântica é realizado o *mashup* dos conceitos identificados e a interação com a camada de Web de Dados, onde é utilizado o DBpedia como fonte de dados. O retorno da anotação semântica é formado por triplas RDF que são armazenadas na camada Servidor, pelo módulo de Armazenamento.



Figura 8 - Marcação de conceitos identificados no metadado descrição.

Após o término do processamento do objeto de aprendizagem, o cliente do Moodle da camada de Aplicação recebe um URI do recurso armazenado e passa a ser disponível para utilização.

Todos os recursos armazenados ficam disponíveis para consulta por meio de uma interface específica, possibilitando assim o reuso dos recursos armazenados. Esta interface realiza as buscas nos repositório em RDF através de consultas SPARQL.

4 CONCLUSÃO

Na educação à distância, disponibilizar materiais de aprendizagem relevantes e de valor com foco no reaproveitamento e a interoperabilidade entre as diversas plataformas de produção de recursos didáticos torna-se um desafio. Portanto, este trabalho apresenta a utilização de técnicas de mineração de textos e uma arquitetura de *web services* para anotação semântica de conteúdo que enriquecem os objetos de aprendizagem, possibilitando novas formas de descoberta de conhecimento e reuso de informação, construindo novas formas de acesso a dados disponíveis em ambientes de aprendizagem, por meio da descrição estruturada em RDF dos objetos de aprendizagem. Além disso, interliga conceitos à fonte de dados na Web de Dados, enriquecendo a base de conhecimento do ambiente de aprendizagem com *mashup* semântico entre estes conceitos e recursos pré-existentes da Web de dados (*Web of Linked Data*).

Por fim, possibilita, um aumento da produtividade no processo de ensino-aprendizagem, uma vez que o usuário, quer seja aprendiz, quer seja instrutor, passa a ser assistido pela máquina, de forma efetiva, na tarefa periférica de busca, integração e inferência ontológica de conhecimento, podendo, pois, focar por mais tempo na tarefa que realmente importa, que é a absorção do referido conhecimento.

5 REFERÊNCIAS BIBLIOGRÁFICAS

ARAUJO, Moysés de. UNIVERSIDADE DE SÃO PAULO, Escola Politécnica. Educação à distância e a WEB Semântica: modelagem ontológica de materiais e objetos de aprendizagem para a plataforma, 2003. 178f, il. Tese (Doutorado).

BECHHOFFER, S.; HARMELEN, F.; HENDLER, J.; HORROCKS, I.; MCGUINNESS, D.; SCHNEIDER, P.; STEIN, L. *OWL Web Ontology Language Reference*. 2004. Disponível em <<http://www.w3.org/TR/owl-ref/>>. Acesso em: 27 nov. 2013.

BERNERS-LEE, T. *Linked Data. Design Issues about Web architecture*. Disponível em <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 15 mai. 2013.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. *The Semantic Web. Scientific American*, Estados Unidos, v.284, p. 34-43, 2001.

CYGANIAK, R.; JENTZSCH, A. *The Linking Open Data cloud diagram*. 2011. Disponível em: <<http://lod-cloud.net/>>. Acesso em: 15 mar. 2014.

GOMES, Geórgia Regina Rodrigues; PONTÍFICA UNIVERSIDADE CATÓLICA. Integração de Repositórios de Sistemas de Bibliotecas Digitais e de Sistemas de Aprendizagem, 2006. 143 p, il. Tese (Doutorado).

GLONVEZYNSKI, Régis Alessandro. UNIVERSIDADE FEDERAL DE SANTA CATARINA. Modelo de anotação de documentos para a codificação do conteúdo semântico no processo de autoria, 2008. Dissertação (Mestrado).

HAWKE, S.; HERMAN, I.; ARCHER, P.; PRUD'HOMMEAUX, E. *Semantic Web Activity - Latest layercake diagram*. 2013. Disponível em: <<http://www.w3.org/2001/sw/>>. Acesso em: 13 abr. 2014.

KAMALELDIN, M.; DUMINDA, W. *Performance Analysis of Web Services on Mobile Devices. Procedia Computer Science*, v. 10, p. 744-751, 2012.

LOPES, Lucelene. PONTÍFICA UNIVERSIDADE CATÓLICA. Extração automática de conceitos a partir de textos em língua portuguesa, 2011. 156p, il. Tese (Doutorado).

LOPES, Maria Célia Santos. UNIVERSIDADE FEDERAL DO RIO DE JANEIRO. Mineração de Dados Textuais Utilizando Técnicas de *Clustering* para o Idioma Português, 2004. 180p. il. Tese (Doutorado).

MONTEIRO, L. de O; GOMES, I. R; OLIVEIRA, T. Etapas do Processo de Mineração de Textos – uma abordagem aplicada a textos em Português do Brasil. Anais: XXVI - Congresso da Sociedade Brasileira de Computação. Campo Grande: UFMS e UCDB, 2006.



MOURA, Simone Leal de. PONTÍFICA UNIVERSIDADE CATÓLICA. Uma arquitetura para Integração de Repositórios de Objetos de Aprendizagem baseada em Mediadores e Serviços Web, 2005. 158 p, il. Dissertação (Mestrado).

OREN, E. *What are Semantic Annotations?* 2006. Disponível em: <<http://www.siegfried-handschuh.net/pub/2006/whatissemannot2006.pdf>>. Acesso em: 05 mar. 2014.

PICKLER, M. E. V; Web Semântica: Ontologias como ferramentas de representação do conhecimento. Perspectivas em Ciência da Informação. *The Scientific Electronic Library Online* – SciELO. 2006.

POPOV, B.; KIRYAKOV, A.; MANOV, D.; KIRILOV, A.; OGNJANOFF, D.; GORANOV, M.; *Towards Semantic Web Information Extraction. Proceedings: II International Semantic Web Conference. Sanibel Island, Florida: 2003.* Disponível em: <<http://gate.ac.uk/conferences/iswc2003/proceedings/popov.pdf>>. Acesso em: 2 abr. 2014.

SOUZA, Renato Rocha; ALVARENGA, Lídia. A Web Semântica e suas contribuições para a ciência da informação. *Ci. Inf.*, Brasília, v. 33, n. 1, p. 132-141, jan./abr. 2004.

SEMANTIC ANNOTATION CONTENT IN SUPPORTING THE PRODUCTION OF LEARNING OBJECTS

Abstract: *In Distance Education, one of the vital elements is the provision of relevant and value learning materials. In recent years, many educational applications Web-based have been developed, but there are still some challenges, among which the search for material and learning objects more intelligent and efficient. For this, one can utilize the concepts introduced with the Semantic Web, where machine- understandable structured metadata are added to the web in order to add meaning to the information. The aggregation of semantic value to documents is a way to organize the publishing process , recovery and enrichment of information , considering that enables the retrieval of information contained in the documents so accurate and effective , facilitating their reuse . Therefore, in this paper is presented a prototype based on a web services architecture for semantic annotation of content, which automates the extraction of concepts in learning objects, using text mining techniques with case study applied to the Moodle learning environment. This work is expected to build new forms of knowledge discovery and reuse of information from data available in learning environments, by describing structured in RDF of learning objects. Furthermore, linking concepts to the data source in the Web of Linked Data, enriching the knowledge base of the learning environment with semantic mashup between these concepts and pre - existing features of Web data, enabling increased productivity in the teaching- learning process.*

Key-words: *Semantic annotation, Learning Object, Linked Data, Distance Education.*