

## UMA FERRAMENTA PARA ENRIQUECER CONTEÚDOS USANDO DADOS ABERTOS CONECTADOS

**LÚCIO LEAL BASTOS<sup>1</sup>; CRISTIAN CECHINEL<sup>2</sup>, DANIELA FRANCISCO BRAUNER<sup>3</sup>;**

<sup>1</sup>Universidade Federal de Pelotas – llbastos@inf.ufpel.edu.br

<sup>2</sup>Universidade Federal de Pelotas – contato@cristiancechinel.pro.br

<sup>3</sup>Universidade Federal de Pelotas – danibrauner@inf.ufpel.edu.br

### 1. INTRODUÇÃO

Em ambientes virtuais de aprendizagem (AVAs), como o Moodle, é comum a disponibilização de inúmeros objetos de aprendizagem em um curso ou módulo. Um objeto de aprendizagem (OA) é qualquer entidade, digital ou não, que pode ser utilizada para aprendizagem, educação ou treinamento (DIAZ-AVILES et al., 2011). No contexto deste artigo, utilizaremos o conceito de OAs como arquivos digitais utilizados como recursos educacionais de cursos a distância.

O crescente volume de dados disponíveis na Web acentua o desafio de busca e recuperação de conteúdo relevante à pesquisa realizada pelo usuário final. Neste sentido, o projeto Linked Open Data, uma iniciativa que visa enriquecer a informação disponibilizada na Web adicionando metadados e *links* entre os dados, facilitando a conexão das informações e enriquecendo ainda mais o conteúdo da Web atual. Dentre esses conjuntos de dados abertos, destaca-se a DBpedia<sup>1</sup>, uma base de dados que disponibiliza conteúdo estruturado relacionado ao conteúdo da Wikipédia. Inúmeras iniciativas e bases de dados disponíveis na Web conectam suas informações ao conteúdo da DBpedia, configurando uma grande base de dados conectados (LOD-CLOUD, 2015).

Diversas ferramentas utilizam-se da base de dados da DBpedia para extração de conteúdo relevante, para diversas finalidades. Dentre estas ferramentas, podemos destacar o DBpedia Spotlight<sup>2</sup>. De acordo com MENDES et al. (2011), o DBpedia Spotlight é uma ferramenta que automaticamente anota menções a recursos da DBpedia em texto, provendo uma solução para ligar dados não estruturados com a base de dados estruturada da DBpedia. O DBpedia Spotlight reconhece nomes de conceitos ou entidades que foram mencionadas e subsequentemente casa esses termos com identificadores únicos (URLs). Pode ser usado para reconhecimento de nomes de entidades, extração de palavras chaves, rotulação, entre outros serviços. A aplicação pode ser acessada através de uma camada de serviço Web, recebendo como entrada um texto e retornando como resultado um HTML anotado com o texto passado para os recursos da DBpedia.

Neste contexto, este trabalho visa apresentar um aplicativo que enriquece os conteúdos de textos de objetos de aprendizagem a partir da sugestão de links para artigos na Wikipédia, utilizando para isso o serviço Web disponibilizado pela DBpedia Spotlight. Os links coletados poderão ser sugeridos aos alunos como referências adicionais ao conteúdo original do curso, enriquecendo o conteúdo disponibilizado pelo professor.

---

<sup>1</sup> <http://wiki.dbpedia.org>

<sup>2</sup> <http://spotlight.dbpedia.org>

## 2. METODOLOGIA

A metodologia utilizada para alcance dos resultados deste trabalho baseou-se nas seguintes etapas: revisão bibliográfica sobre linked data e as ferramentas propostas, implementação e testes. Essa metodologia, foi sendo realizada de forma interativa e incremental, de forma que, a cada novo ciclo, incrementalmente a aplicação foi sendo melhorada. Por exemplo, os testes foram alimentando melhorias no desenvolvimento da aplicação.

## 3. RESULTADOS E DISCUSSÃO

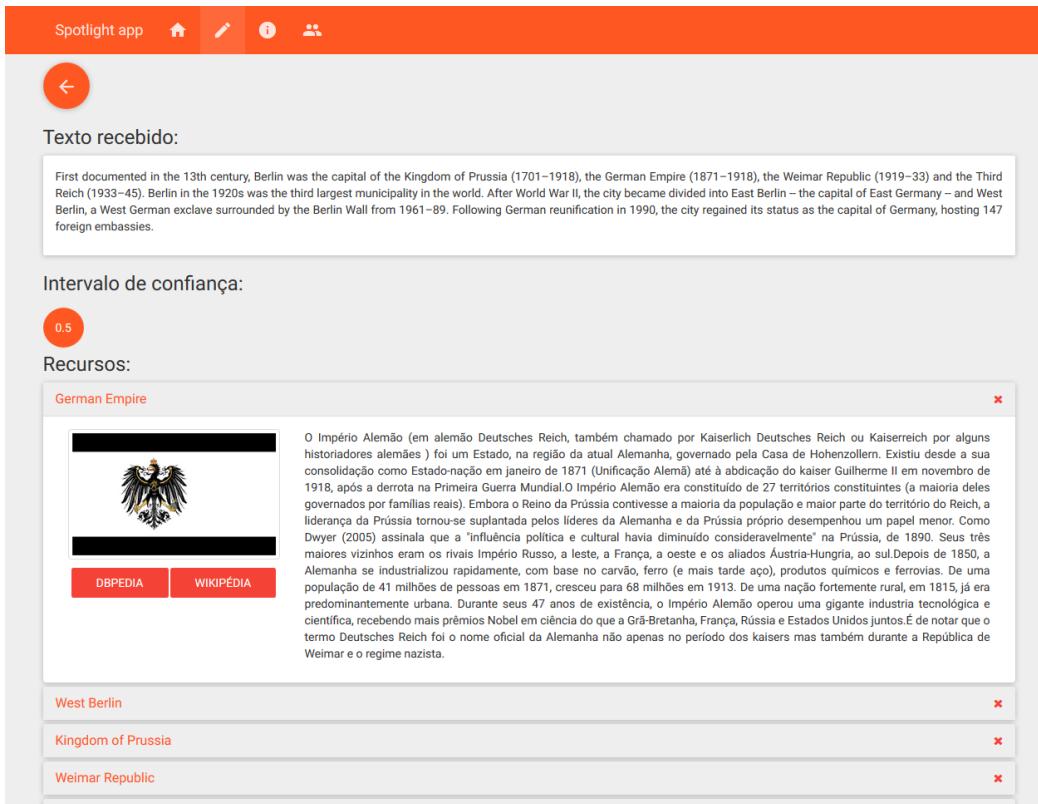
Como resultado foi desenvolvido o Spotlight App, uma aplicação que recebe como entrada um arquivo em texto e realiza conexão com um servidor da DBpedia, o qual através de suas informações estruturadas por meio de ontologias, retorna os dados relevantes que serão tratados e devolvidos à aplicação. Esta aplicação tem o objetivo de realizar marcações de palavras-chave, filtradas através de mecanismos internos ao DBpedia Spotlight, respeitando um nível de relevância em relação ao contexto da frase em que se encontra.

Para a consulta na base de dados foi utilizado um *endpoint* do DBpedia Spotlight, passando como consulta o texto de entrada (vide Figura 1) e obtendo como resultado uma página Web (vide Figura 2) que exibe os termos anotados, um resumo de cada termo, uma imagem em miniatura e o link para a Wikipédia.

A aplicação foi desenvolvida em *Javascript* com auxílio do *framework* *jQuery*, e apresentado com *HTML* e *CSS*, com auxílio do *framework* *Bootstrap*, utilizando o tema Material Design. As respostas do DBpedia Spotlight foram dadas no formato *JSON* (*JavaScript Object Notation*), escolhido pela simplicidade em lidar com ele em *Javascript*.

```
SELECT ?resumo ?imagem ?pt
WHERE {
    <http://dbpedia.org/resource/Universidade_Federal_de_Pelotas>
    <http://dbpedia.org/ontology/abstract> ?resumo;
    <http://dbpedia.org/ontology/thumbnaill> ?imagem;
    <http://www.w3.org/2002/07/owl#sameAs> ?pt.
    FILTER (contains(xsd:string(?pt), "pt.dbpedia.org") && langMatches(lang(?resumo), "pt"))
}
```

Figura 1: Consulta SPARQL que retorna o resumo, imagem e URI do recurso em português.



The screenshot shows the Spotlight App interface. At the top, there is a toolbar with icons for 'Spotlight app', a home icon, a pencil, a magnifying glass, a person, and a user profile. Below the toolbar, there is a back arrow button. The main area displays the text received:

First documented in the 13th century, Berlin was the capital of the Kingdom of Prussia (1701–1918), the German Empire (1871–1918), the Weimar Republic (1919–33) and the Third Reich (1933–45). Berlin in the 1920s was the third largest municipality in the world. After World War II, the city became divided into East Berlin – the capital of East Germany – and West Berlin, a West German exclave surrounded by the Berlin Wall from 1961–89. Following German reunification in 1990, the city regained its status as the capital of Germany, hosting 147 foreign embassies.

Below the text, there is a section for 'Intervalo de confiança:' with a value of '0.5'. Under 'Recursos:', there is a card for 'German Empire' which includes a thumbnail of the German eagle logo, two buttons for 'DBPEDIA' and 'WIKIPÉDIA', and a detailed description of the German Empire's history. Below this, there are cards for 'West Berlin', 'Kingdom of Prussia', and 'Weimar Republic', each with a red 'X' icon in the top right corner.

Figura 2: Interface da Spotlight App contendo como resultado os links extraídos do texto de entrada.

## 4. CONCLUSÕES

Este trabalho apresentou uma ferramenta para enriquecer conteúdos usando dados abertos conectados. Alguns testes iniciais, durante o desenvolvimento da aplicação, foram desenvolvidos demonstrando a funcionalidade de extrair links relevantes da Wikipedia usando a DBpedia através da DBpedia Spotlight, o que demonstra que a metodologia proposta é adequada para continuação em trabalhos futuros.

Este trabalho é parte do resultado de um trabalho de conclusão de curso, cujo objetivo final é enriquecer objetos de aprendizagem no formato PDF, de forma integrada ao Ambiente Virtual de Aprendizagem (AVA) da Universidade Federal de Pelotas. Em uma primeira iniciativa, será desenvolvido um *plugin* para ser integrado ao AVA, no qual professores possam realizar consultas sobre seus OAs, escolher resultados que considerem relevantes, disponibilizando estes com o OA para seus alunos. Caso não seja possível esta abordagem, serão então desenvolvida extensões para os principais navegadores do mercado, que realizarão esta funcionalidade assim que o navegador detectar que o professor ou aluno entrar no ambiente do AVA.

Alguns pontos importantes devem ser observados sobre a experiência relatada neste artigo, que servirão de referência para os trabalhos futuros:

- **Incompatibilidade com o formato PDF.** O formato de arquivos PDF é largamente utilizado para distribuição de OAs, portanto é interessante seu suporte pela ferramenta. Um detalhe deste formato de arquivo é que sua

estrutura prioriza a apresentação, e não o conteúdo em si, o que pode dificultar a obtenção de parágrafos de forma correta;

- **Instabilidade no endpoint SPARQL.** O endpoint utilizado pelo aplicativo possui momentos de instabilidade, os quais implicam em resultados imprecisos no seu uso. Uma alternativa para resolver este problema é a instalação de um *endpoint* local do DBpedia Spotlight na Universidade Federal de Pelotas, porém o alto requerimento de hardware, até o presente momento, impossibilitou esta alternativa;
- **Enriquecimento dos resultados.** O JSON retornado na consulta pode fornecer outras informações sobre o conteúdo, como mais imagens, coordenadas geográficas, *links*, etc. Uma opção a ser explorada para a melhora da consulta é a seleção de um número variável de atributos a partir da classe da ontologia a qual o recurso pertence;
- **Utilização de outras bases de conhecimento.** Um incremento pertinente para a aplicação é o uso de outras bases de conhecimento em conjunto com a DBpedia, também baseado na classe da ontologia do recurso pesquisado.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

DIAZ-AVILES, E., FISICHELLA, M., KAWASE, R., NEJDL, W., STEWART, A. Unsupervised auto-tagging for learning object enrichment. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 6964, p. 83–96, 2011.

LOD-CLOUD. **The Linking Open Data cloud diagram**. 23 jul. 2015. Acessado em 23 jul. 2015. Online. Disponível em <http://lod-cloud.net>

MENDES, P. N., JAKOB, M., GARCÍA-SILVA, A., BIZER, C. **DBpedia Spotlight : Shedding Light on the Web of Documents**. In **Proceedings of the 7th International Conference on Semantic Systems (I-Semantics)**, v. 95, p. 1–8, 2011.