

O Papel das Bases de Conhecimento Web :

Eduardo Kenji Avena

Universidade de São Paulo - USP, São Paulo-SP, Brasil

edukenjister@gmail.com

## RESUMO

Extraído do conteúdo da Wikipédia, a DBpedia é considerada uma das mais importantes bases de conhecimento da Web Semântica, que possui edições em diversos idiomas, entre os quais o inglês (DBpedia EN) e o português (DBpedia PT). Todas as edições da DBpedia estão sujeitas a problemas de qualidade, mais especialmente a DBpedia PT sofre de inconsistências e falta de dados em vários domínios. Este artigo descreve um processo semi-automático e incremental de publicação de dados na DBpedia, provenientes de fontes externas confiáveis, eficientes e eficazes, buscando aprimorar aspectos de sua qualidade. Em um contexto de ciência aberta, a proposta visa consolidar a DBpedia como um hub de referência para dados de pesquisa, de forma que pesquisas de qualquer área suportada pelos dados da Web Semântica possam utilizar seus dados de forma confiável. Embora a abordagem seja independente de uma edição específica da DBpedia, a ferramenta de protótipo de suporte, chamada ETL4DBpedia, foi construída para a DBpedia PT, com base em fluxos de trabalho ETL (Extract, Transform, Load). Este artigo também descreve a avaliação da abordagem, aplicando a ferramenta em um cenário de uso real envolvendo dados da área de botânica. Esta aplicação resultou em um aumento de 127% na completude de espécies de plantas medicinais na DBpedia PT, além de apresentar desempenho satisfatório para os componentes da ETL4DBpedia.

## Referências

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak e Zachary Ives. 2007. DBpedia: A Nucleus for a Web of Open Data. In Proceedings of the 6th International The Semantic Web and 2nd Asian Conference on Asian Semantic Web Conference (ISWC'07 / ASWC'07). Springer-Verlag, Berlin, Heidelberg, 722--735. Biblioteca Digital do Google Scholar

T. Berners-Lee. 2009. Dados vinculados. Recuperado em 28 de julho de 2020 em <https://www.w3.org/DesignIssues/LinkedData.html> Google Scholar

Daniilo Barbosa Coimbra, José Olímpio Mouzinho Negrão e Frederico Araujo Durão. 2019. LODGVis: uma visualização interativa para navegação de dados abertos vinculados. Em Anais do 25º Simpósio Brasileiro de Multimídia e a Web (WebMedia '19). Association for Computing Machinery, Nova York, NY, EUA, 433--440. <https://doi.org/10.1145/3323503.3360300> Google Scholar Digital Library

Mostrar todas as referências

Termos do Índice

Uma abordagem para melhorar a DBpedia como um hub de dados de pesquisa

Sistemas de informação

Sistemas de gerenciamento de dados

Integração de informação

Extração, transformação e carregamento

Recuperação de informação

Tarefas e objetivos de recuperação

Extração de informação

Rede mundial de computadores

Linguagens de descrição de dados da web

Linguagens de descrição da web semântica

Perguntas:

1)"Qual a principal mensagem do trabalho?" Para responder a esta pergunta, identifique no artigo qual o problema que os autores procuram resolver

2)"Qual a principal contribuição do trabalho?" Para responder a esta pergunta, identifique no artigo qual a solução reportada pelos autores

3)"Como o trabalho foi avaliado?" Para responder a esta pergunta, identifique no artigo como os autores avaliaram o impacto da solução proposta no problema identificado originariamente

4)"Qual a principal limitação do trabalho?" Para a responder esta pergunta, *analise criticamente* o trabalho apresentado no artigo e identifique uma limitação na solução reportada relativamente ao problema tratado

Sua resposta

1) Extraído do conteúdo da Wikipédia, a DBpedia é considerada uma das mais importantes bases de conhecimento da Web Semântica, que possui edições em diversos idiomas do mundo, entre os quais o inglês (DBpedia EN) e o português (DBpedia PT). Todas as edições da DBpedia estão sujeitas a problemas de qualidade, eficiência e eficácia, mais especialmente a DBpedia PT sofre de inconsistências e falta de dados em vários domínios ( saúde, educação, segurança, meio-ambiente, transporte, ciência, tecnologia, esporte, economia, geografia, cidadania, energia, cultura, limpeza urbana, etc...);

2) Este artigo descreve um processo semi-automático, automatizado e incremental de publicação de dados na DBpedia, provenientes de fontes externas confiáveis, eficientes e eficazes, buscando aprimorar aspectos de sua qualidade e quantidade. Em um contexto de ciência aberta, a proposta visa consolidar a DBpedia como um "hub" de referência para dados de pesquisa, de forma que pesquisas de qualquer área suportada pelos dados da Web Semântica possam utilizar seus dados de forma confiável, eficiente e eficaz. Embora a abordagem seja independente de uma edição específica da DBpedia, a ferramenta de protótipo de suporte, chamada ETL4DBpedia, foi construída para a DBpedia PT, com base em fluxos de trabalho ETL (Extract, Transform, Load);

3) Este artigo também descreve a avaliação da abordagem, aplicando a ferramenta em um cenário de uso real envolvendo dados da área de botânica, plantas e vegetações. Esta aplicação resultou em um aumento de 127% na completude de espécies de plantas e ervas medicinais na DBpedia PT, além de apresentar desempenho satisfatório para os componentes da ETL4DBpedia;

4) Limitações quanto a veracidade, origem, duplicidade de informações, falta de informações e dados de certos domínios como ( saúde, geografia, etnia, educação, segurança, meio-ambiente, transporte, ciência, tecnologia, esporte, economia, geografia, cidadania, energia, cultura, limpeza urbana, etc...).