

Title:

DBpedia – An Interlinking Hub connecting the Web of Documents and the Web of Data

Abstract:

DBpedia is a community effort to extract structured information from Wikipedia and to make this information available on the Web. The resulting DBpedia knowledge base currently describes more than 3.4 million things out of which 1.5 million are classified in a consistent ontology. Concepts within the DBpedia knowledge are interlinked with 5,081,000 external web pages and 9,393,000 million records in external Linked Data sources. Within the talk, we will give an overview of the DBpedia project and the emerging Web of Linked Data. Afterwards, we will describe two open-source tools that are currently developed by our group in order to increase the linkage between DBpedia and external Web pages as well as records in external Linked Data sources:

1. The *Silk – Link Discovery Framework*, an identity resolution tool for discovering relationships between records within different Linked Data sources. Silk generates RDF links based on declarative, class-level matching descriptions. Matching jobs can be executed on Hadoop clusters.
2. The *DBpedia Resource Recognizer*, a tool for annotating free-text with DBpedia URIs which uses the DBpedia knowledge base as background knowledge for disambiguation. The tool recognizes surface forms of DBpedia concepts by exploiting the context in which a surface form appears using traditional vector-based models as well as graph-based measures on the DBpedia knowledge base.

By providing more links between DBpedia and external web pages as well as Linked Data sources, we hope to increase the utility of DBpedia as an interlinking hub connecting the classic Web of Documents and the emerging Web of Data and thus to enable applications, like next-generation search engines, to exploit the benefits of both worlds.

The Speakers:

Prof. Dr. Christian Bizer

Chris Bizer is the head of the Web-based Systems Group at Freie Universität Berlin. The group explores technical and economic questions concerning the development of global, decentralized information environments. Christian Bizer initialized the W3C Linking Open Data community project and the DBpedia project, an effort to extract structured data from Wikipedia and to publish the extracted data as Linked Data on the Web. Other results of his work include the Named Graphs data model which was adopted into the W3C SPARQL standard, the D2RQ mapping language which is widely used for publishing relational databases to the Web of Linked Data, and the Berlin SPARQL Benchmark for measuring the performance of RDF stores. More information about Chris is found at <http://www.wiwiss.fu-berlin.de/en/institute/pwo/bizer/team/BizerChristian.html>

MSc Max Jakob

Max Jakob is a member of the Web-based Systems Group at Freie Universität Berlin since June 2010. He acquired his Master of Science degree in Language & Communication Technologies from both Saarland University and Charles University in Prague, focusing on different semantic representations

for natural language processing. His current projects are the DBpedia Resource Recognizer and the maintenance of the DBpedia extraction framework. More information about Max is found at <http://www.wiwiss.fu-berlin.de/en/institute/pwo/bizer/team/JakobMax.html>

References:

Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, Sebastian Hellmann: DBpedia - A crystallization point for the Web of Data. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, Issue 7, Pages 154–165, 2009.

Christian Bizer, Tom Heath, Tim Berners-Lee: Linked Data - The Story So Far. In: *International Journal on Semantic Web & Information Systems*, Vol. 5, Issue 3, Pages 1-22, 2009.

Julius Volz, Christian Bizer, Martin Gaedke, Georgi Kobilarov: Discovering and Maintaining Links on the Web of Data. *International Semantic Web Conference (ISWC2009)*, Westfields, USA, October 2009.