Linked Data on the Web (LDOW2008)

Workshop at the 17th International World Wide Web Conference
Beijing, China, April 22, 2008

Christian Bizer	Tom Heath	Kingsley Idehen	Tim Berners-Lee
Freie Universität Berlin	Talis	OpenLink Software	W3C
Germany	United Kingdom	USA	USA
chris@bizer.de	tom.heath@talis.com	kidehen@openlinksw.com	timbl@w3.org

ABSTRACT

The Web is increasingly understood as a global information space consisting not just of linked documents, but also of Linked Data. More than just a vision, the resulting Web of Data has been brought into being by the maturing of the Semantic Web technology stack, and by the publication of an increasing number of data sets according to the principles of Linked Data.

The Linked Data on the Web (LDOW2008) workshop brings together researchers and practitioners working on all aspects of Linked Data. The workshop provides a forum to present the state of the art in the field and to discuss ongoing and future research challenges. In this workshop summary we will outline the technical context in which Linked Data is situated, describe developments in the past year through initiatives such as the Linking Open Data community project, and look ahead to the workshop itself.

Categories and Subject Descriptors

H.3.5. [Information Storage and Retrieval]: Online Information Services – Data sharing. H.3.5. [Information Storage and Retrieval]: Online Information Services – Web-based services.

1. THE WEB OF DATA

Linked Data is about employing the Resource Description Framework (RDF) and the Hypertext Transfer Protocol (HTTP) to publish structured data on the Web and to connect data between different data sources, effectively allowing data in one data source to be linked to data in another data source.

The principles of Linked Data were first outlined by Berners-Lee in 2006 [1], and provide broad guidance upon which data publishers have begun to realize the Web of Data. This guidance has been extended by technical documents such as [2][3] that capture best practices emerging from the Linked Data community and provide recipes on which publishing systems can be based.

The Web of Data can be accessed using Linked Data browsers, just as the traditional Web of documents is accessed using HTML browsers. However, instead of following links between HTML pages, Linked Data browsers enable users to navigate between different data sources by following RDF links. This allows the user to start with one data source and then move through a potentially endless Web of data sources connected by RDF links. Just as the traditional document Web can be crawled by following

hypertext links, the Web of Data can be crawled by following RDF links. Working on the crawled data, search engines can provide sophisticated query capabilities, similar to those provided by conventional relational databases. Because the query results themselves are structured data, not just links to HTML pages, they can be immediately processed, thus enabling a new class of applications based on the Web of Data.

The glue that holds together the traditional document Web is the hypertext links between HTML pages. The glue of the data web is RDF links. An RDF link simply states that one piece of data has some kind of relationship to another piece of data. These relationships can have different types. For instance, an RDF link that connects data about people can state that two people know each other; an RDF link that connects information about a person with information about publications in a bibliographic database might state that a person is the author of a specific paper.

2. THE LINKING OPEN DATA PROJECT

Evidence of the emergence of a Web of Data comes from the Linking Open Data project[1], a grassroots community effort founded in February 2007 and supported by the W3C Semantic Web Education and Outreach Working Group[2]. The aim of the project is to identify data sets that are available under open licenses, re-publish these in RDF on the Web and interlink them with each other.

During 2007, the size of the Web of Data that directly stems from this community effort has grown to over two billion RDF triples, originating from data sets in diverse domains such as geographic information, census information, people, companies, online communities, human languages, scientific publications, films, music, books and reviews[3]. These data sets are interlinked by around three million RDF links.

An indication of the range and scale of the Linking Open Data "cloud" is provided in Figure 1. As this diagram shows, key interlinking hubs are sites such as DBpedia[4] and Geonames[5]. DBpedia extracts RDF triples from the "Infoboxes" commonly seen on the right hand side of Wikipedia articles, and makes these available on the Web in RDF to be crawled or queried with

[1] http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/ LinkingOpenData

[2] http://www.w3.org/2001/sw/sweo/

[3] http://richard.cyganiak.de/2007/10/lod/

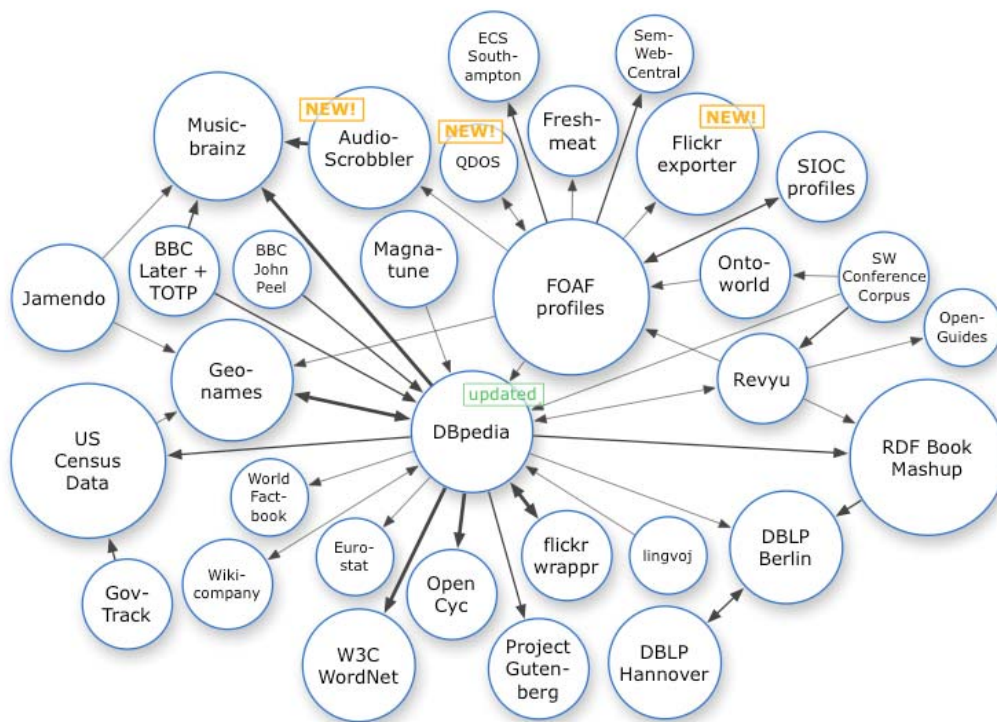[4] http://dbpedia.org/About

[5] http://www.geonames.org/

Figure 1. The Linking Open Data "cloud", February 2008

SPARQL. Geonames in turn provides RDF descriptions of millions of geographical locations worldwide. DBpedia and Geo-names provide URIs (and RDF descriptions) for many of the things in the world to which we want to refer. As these URIs are reused within other data sets, DBpedia and Geonames develop into hubs to which an increasing number of other data sets are connected, thereby increasing the potential for network effects.

In addition to publishing and interlinking data sets, there is also ongoing work within the community project on Linked Data browsers, Linked Data crawlers, Web of Data search engines and other applications that consume Linked Data from the Web. Examples of Linked Data browsers[6] are Tabulator, Disco, the OpenLink data browser and the Zitgist browser. Examples of Web of Data search engines[7] and lookup indexes are Falcons, Sindice, Swoogle and Watson. These services enable humans and machines to locate and query Linked Data that has been published across the Web.

3. THE LDOW2008 WORKSHOP

The goal of the workshop is to provide a forum for the Linked Data community, in which participants can present and discuss approaches to publishing Linked Data on the Web, showcase innovative applications that consume Linked Data, and shape the agenda and identify upcoming research issues for the next development stage of the Web of Data.

The program of papers to be presented at the workshop demonstrates ongoing innovation not just in populating the Web of data, but in developing applications that exploit this ecosystem for end users, and in advancing conceptual and architectural issues related to the field. In total 37 papers and demos were submitted for review, of which 23 were accepted for presentation at the workshop and inclusion in the proceedings.

We would like to extend our sincere thanks to the LDOW2008 Program Committee for the efforts they expended in reviewing the submitted papers, and to the WWW2008 organizers for their support of the workshop. For more information about the workshop program please refer to the workshop website:

http://events.linkeddata.org/ldow2008/

4. REFERENCES
[1] Berners-Lee, T. Linked Data. 2006. http://www.w3.org/DesignIssues/LinkedData.html

[2] Bizer, C., Cyganiak, R., Heath, T. How to Publish Linked Data on the Web. 2007. http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/

[3] Sauermann, L., Cyganiak, R., Ayers, D., Völkel, M. Cool URIs for the Semantic Web. 2007. http://www.w3.org/TR/cooluris/

---

[6]http://esw.w3.org/topic/TaskForces/CommunityProjects/ LinkingOpenData/SemWebClients

[7]http://esw.w3.org/topic/TaskForces/CommunityProjects/ LinkingOpenData/SemanticWebSearchEngines