Linked Data - The Story So Far

Christian Bizer, Freie Universität Berlin, Germany
Tom Heath, Talis Information Ltd, United Kingdom
Tim Berners-Lee, Massachusetts Institute of Technology, USA

This is a preprint of a paper to appear in: Heath, T., Hepp, M., and Bizer, C. (eds.). Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS). http://linkeddata.org/docs/ijswis-special-issue

Abstract

The term Linked Data refers to a set of best practices for publishing and connecting structured data on the Web. These best practices have been adopted by an increasing number of data providers over the last three years, leading to the creation of a global data space containing billions of assertions - the Web of Data. In this article we present the concept and technical principles of Linked Data, and situate these within the broader context of related technological developments. We describe progress to date in publishing Linked Data on the Web, review applications that have been developed to exploit the Web of Data, and map out a research agenda for the Linked Data community as it moves forward.

Keywords: Linked Data, Web of Data, Semantic Web, Data Sharing, Data Exploration

1. Introduction

The World Wide Web has radically altered the way we share knowledge by lowering the barrier to publishing and accessing documents as part of a global information space. Hypertext links allow users to traverse this information space using Web browsers, while search engines index the documents and analyse the structure of links between them to infer potential relevance to users' search queries (Brin & Page, 1998). This functionality has been enabled by the generic, open and extensible nature of the Web (Jacobs & Walsh, 2004), which is also seen as a key feature in the Web's unconstrained growth.

Despite the inarguable benefits the Web provides, until recently the same principles that enabled the Web of documents to flourish have not been applied to data. Traditionally, data published on the Web has been made available as raw dumps in formats such as CSV or XML, or marked up as HTML tables, sacrificing much of its structure and semantics. In the conventional hypertext Web, the nature of the relationship between two linked documents is implicit, as the data format, i.e. HTML, is not sufficiently expressive to enable individual entities described in a particular document to be connected by typed links to related entities.

However, in recent years the Web has evolved from a global information space of linked documents to one where both documents and data are linked. Underpinning this evolution is a set of best practices for publishing and connecting structured data on the Web known as