Chapter 13
Linked Data, Towards Realizing the Web of Data: An Overview

Leila Zemmouchi-Ghomari
University of Sciences and Technology Houari Boumediene (USTHB), Algeria

ABSTRACT

Data play a central role in the effectiveness and efficiency of web applications, such as the Semantic Web. However, data are distributed across a very large number of online sources, due to which a significant effort is needed to integrate this data for its proper utilization. A promising solution to this issue is the linked data initiative, which is based on four principles related to publishing web data and facilitating interlinked and structured online data rather than the existing web of documents. The basic ideas, techniques, and applications of the linked data initiative are surveyed in this paper. The authors discuss some Linked Data open issues and potential tracks to address these pending questions.

INTRODUCTION

At present, the web is based on several notions of information sharing (Allemang and Hendler, 2011) that lead to considerable misunderstanding, such as the Anyone can say Anything about Any topic (AAA) slogan, the open world assumption, and the non-unique naming assumption. Thus, there is a pressing need to move from the current framework to one guided by consistent (homogeneous) principles that allow information sharing, cooperation, and collaboration.

Web data often consists of isolated silos (Herman, 2010) that cannot exchange content with other systems on the web. This lack of communication and sharing of data is due to incompatibility among the various online data formats. A contextual interpretation of these datasets is time-consuming and expensive, and it requires developers' intervention. Moreover, data is embedded into web pages and is available only for human consumption. In contrast, the Semantic Web aims to enable machines to understand and process data contained in web pages and online documents. It targets the realization of a structured 'Web of Data' to complement the existing Web of Documents, which is loosely structured at best.

DOI: 10.4018/978-1-5225-5191-1.ch013

Many researchers (Bizer & Heath, 2009; Wood et al., 2014; Hogan, 2014) are convinced that the Linked Data initiative[1] is a promising approach for publishing and connecting structured data on the Internet by means of a set of standards and tools. Linked Data technologies render explicit and transparent the conceptual models underlying the visible data. In particular, they support data integration in dynamic and distributed environments, such as large enterprises, intergovernmental organizations, and the World Wide Web (Mendez & Greenberg, 2012).

Data exchange is possible if everyone uses a shared data format, such as the Resource Description Framework (RDF), as well as common ways of accessing it, such as Simple Protocol and RDF Query Language (SPARQL).

In practice, Linked Data facilitates the creation of web pages using information from multiple web pages. As Wood notes, 'Linked Data enables cooperation without coordination' (Wood et al., 2014).

However, simply publishing Linked Data in the Linked Data cloud does not allow reuse. Publishing requires provenance, quality, credit, attribution, and the implementation of methods to provide reproducibility for the validation of results (Bechhofer et al., 2013). Thus, several challenges remain unaddressed for researchers in this field.

Survey research can be used for exploration, description, or explanation purposes (Pinsonneault & Kraemer, 1993). Survey research in exploration aims to make readers more familiar with a topic and explain its basic concepts. The objective of this paper is to highlight the growing movement of the Semantic Web community towards the realization of the Web of Data using Linked Data principles for publishing and consuming data on the web. We explore the best practices and recommendations of the World Wide Web Consortium (W3C) in terms of web data modelling and querying, interlinking techniques, vocabulary reuse, and key areas for future research.

The remainder of this paper is organized as follows. Section 2 explains the main Web of Data principles with regard to the Web of Documents and describes how this can be realized through different approaches. Section 3 defines Linked Data and states its four principles (Section 3). Section 4 describes RDF, the Linked Data model. Section 5 provides an overview of SPARQL, the Linked Data query language. Sections 6 and 7 discuss existing approaches and tools for the publication and consumption of Linked Data from heterogeneous data sources and formats. Section 8 presents some of the most prominent success stories of the application of Linked Data technologies. Section 9 discusses the challenges faced by Linked Data along with the probable causes and possible solutions. Finally, Section 10 concludes the paper by highlighting the importance of Linked Data as a promising research field.

WEB OF DATA DESCRIPTION AND APPROACHES

The Web of Data (Bizer & Heath, 2011) can be considered as another layer that is linked with the classic document Web, and it has the following features:

The Web of Data is generic and can contain any type of data.

Entities are connected by links, creating a unique giant global graph that extends data sources and allows new data sources to be found.

The Web of Data is open, meaning that applications do not have to be implemented against a fixed set of data sources, and anyone can publish data to the Web.

In order to facilitate a better understanding of the Web of Data, a comparison between the Web of Data and the Web of Documents is presented in Table 1.