Chapter 92
The Use of Software Tools in Linked Data Publication and Consumption:
A Systematic Literature Review

Armando Barbosa
Federal University of Alagoas, Maceió, Brazil

Ig Ibert Bittencourt
Federal University of Alagoas, Maceió, Brazil

Sean Wolfgand Matsui Siqueira
Federal University of the State of Rio de Janeiro (UNIRIO), Rio de Janeiro, Brazil

Rafael de Amorim Silva
Federal University of Alagoas, Maceió, Brazil

Ivo Calado
Federal Institute of Alagoas, Rio Largo, Brazil

ABSTRACT

To reduce the complexity intrinsic to LD manipulation, software tools are used to publish or consume data associated to LD activities. However, few developers have a broad understanding of how software tools may be used in publication or consumption of Linked Data. The goal of this work is to investigate the use of software tools in Linked Data publication and consumption processes. More specifically, understanding how these software tools are related to process of publication or consumption of LD. In order to meet their goal, the authors conducted a Systematic Literature Review (SLR) to identify the studies on the use of software tools in these processes. The SLR gathered 6473 studies, of which only 80 studies remained for final analysis (1.25% of the original sample). The highlights of the study are: (1) initial steps of the publication process are fairly supported by the software tools; (2) Non-RDF serialization is fairly supported in publication and consumptions process by the software tools; and (3) there are non-supported steps in consumption and publication processes by the tools.

DOI: 10.4018/978-1-7998-7297-9.ch092

1. INTRODUCTION

In last decade, the evolution of the Web has increased the consumption and production of content. This growth implies in several challenges such as a longer time required to perform searches or browsing, data consistency problems, and misinterpretations. The users and machines perceive these challenges, as they are necessary to access a set of web pages to find relevant information. In addition, most of the data available on the Web are unstructured (images, blogs, web content, etc.), which means that machines are not able to automatically understand such data content. Therefore, it is essential to provide alternatives for structuring data in a machine understandable manner.

Linked Data (LD) is an effective option to structure such data, presenting a set of best practices for guiding publishing and connecting these structured data on the Web (Hyland & Wood, 2011). LD adopts a generic abstract data model for describing resources called Resource Description Framework (RDF). This framework is responsible for describing terms and their relationships, by using triples of subjects, predicates and objects. Additionally, ontologies may provide enrichment to Linked Data by adding semantics to their RDF statements (Heath & Bizer, 2011). Moreover, LD brings some benefits which may be applied in multiple research fields, such as data interoperability and terms definition (Bauer & Kaltenböck, 2011). Some applications of LD include Open Government (Ding, Lebo, Erickson, DiFranzo, Williams, Li, & Hendler, 2011; Shadbolt, O'Hara, Berners-Lee, Gibbins, Glaser, & Hall, 2012) (e.g. to improve of transparency and provenance), Education (Dietze, Yu, Giordano, Kaldoudi, Dovrolis, & Taibi, 2012) (e.g. to improve education repositories and content sharing), and Business (Barnes & Martin, 2002) (e.g. to improve business opportunities and business value).

LD is defined by both publication and consumption processes. Recently, W3C (2014) has published a list of 10 steps to fulfill LD publication requirements: (1) to prepare stakeholders; (2) to select dataset for reuse; (3) to model the data to represent data objects and their relationships; (4) to specify a suitable open data license; (5) to define URI for linked data; (6) to describe objects with standard vocabularies; (7) to convert data to a linked data representation; (8) to provide machine access to data; (9) to announce to the public; and (10) to establish a social contract of a linked data publisher. On the other hand, Bauer and Kaltenböck (2011) have published a list of 7 steps related to LD consumption process: (1) to specify concrete use cases for new services or applications; (2) to evaluate relevant data sources and datasets; (3) to check licenses for use and reuse provided by data owners; (4) to create consumption patterns to specify what data are reused from a certain data source; (5) to manage alignment, caching and update mechanisms; (6) to create mashups and applications to provide user-friendly graphical user interfaces (GUIs) and powerful services for end users; and (7) to establish sustainable partnerships.

Both aforementioned processes are essentials to improve data quality, interoperability, and discoverability of data, as well as to support the development of rich applications. Additionally, considering the number of steps and activities, it is important the development of tools for supporting Linked Data publication and consumption processes. Considering the huge number of Linked Data-related tools proposed recently (Bizer, Heath, & Berners-Lee, 2009), it is a time-consuming task, for publishers or consumers, to know the available features of each tool in order to make properly decision about software tools that should be used according to each step. For this reason, a high-level question was raised: Which software tools have been used to support the publication and/or consumption of Linked Data? In this context, our purpose is to identify what tools have been used to perform the aforementioned tasks, providing a broad vision about Linked Data software tools to publish and/or consume linked data. Additionally, we use the categories provided by W3C[1] to classify the software tools according to their features.