On the Current State of Linked Open Data: Issues, Challenges, and Future Directions

Nosheen Fayyaz, University of Peshawar, Peshawar, Pakistan
Irfan Ullah, University of Peshawar, Peshawar, Pakistan
Shah Khusro, University of Peshawar, Peshawar, Pakistan

ABSTRACT

This article describes how Linked Open Data (LOD), under the umbrella of the Semantic Web, integrates the openly-published semantic information making it easily understandable and consumable by humans and machines. Currently, researchers have applied the principles of LOD in several domains including e-government, media, publications, geography, and life sciences. Besides the fast pace of research, the field is still an emerging one, where researchers face several prominent challenges and issues that need to resolve to exploit LOD to its fullest. In this article, the authors have identified challenges, issues, and research opportunities in the publishing, management, linking, and consumption of LOD. The research work presented here will grab the attention of researchers and may aid to the current state-of-the-art in this area.

KEYWORDS

Big Data, Information Retrieval, Linked Data, Linked Open Data, Ontology, Semantic Web

INTRODUCTION

The World Wide Web is growing at a fast pace with lots of data available, but finding relevant information efficiently and timely is still a challenging job. Therefore, a layer of semantics is added to the current Web to make the machines understand the web documents (Berners-Lee, Hendler, & Lassila, 2001). The Semantic Web not only puts data on the Web but also establishes links among data so that humans and machines can explore the web of data rather than the web of documents and find other related data (Berners-Lee, 2006; Bizer, Heath, & Berners-Lee, 2009). This makes the Web more meaningful with the emergence of new paradigms such as Linked Open Data (LOD). Linked Open Data, an application of Semantic Web, can be defined as the Linked Data that is released under the open license and can be reused for free. Tim Berners-Lee presented a set of four basic rules for publishing, connecting and consuming the structured data on the Web (Berners-Lee, 2006). This Linked Data should have some general features including openness, modularity, scalability, and connectivity (Berners-Lee, 2009). These features make LOD as a source of knowledge discovery with several applications in multiple domains including e-government, health, education, and research.

Publishing data on LOD requires five basic steps including data preparation, dataset selection, plugging into the LOD cloud, its announcement, and identifying social contract (“Best Practices for Publishing Linked Data,” 2014; “Cookbook for Open Government Linked Data,” 2011; Hausenblas, 2009; Peinl, 2016). Figure 1 graphically illustrates these steps. First, the data is converted to triples. Second, the desired datasets (which could enrich the content) are selected (using the “follow your
nose” principle). Third, the dataset is plugged into the LOD cloud for integrating data seamlessly to the chosen datasets to allow users and machines to consume it. Fourth, the data is published on a web server for its use and reuse. Finally, the implicit social contract with the consumers is recognized, which includes the responsibility of data availability, data freshness, accuracy, and maintenance (“Best Practices for Publishing Linked Data,” 2014; “Cookbook for Open Government Linked Data,” 2011; Hausenblas, 2009). The details of tools for the data extraction, transformation, storage, visualization and quality in publishing data on LOD can be found in the relevant literature (Bikakis, Tsinaraki, Gioldasis, Stavrakantonakis, & Christodoulakis, 2013; Butt, Haller, & Xie, 2015; Ferrara, Meo, Fiumara, & Baumgartner, 2014; Gangemi, 2013; Purohit et al., 2016).

Several review articles and survey papers have been published on the various aspects of the state-of-the-art in Linked Open Data and its applications. (Purohit et al., 2016) identified the best practices for publishing scientific research Linked Data and explored several tools for data extraction, transformation, storage, visualization and quality. Similarly, (Mouzakitis et al., 2017) explored the state-of-the-art in Linked Data technologies especially tools and identified their shortcomings and common barriers in their usage. They also proposed a framework to maintain and promote the interlinking capabilities of Linked Data so that novice users may use it (Mouzakitis et al., 2017). (Emani, Cullot, & Nicolle, 2015) focused big Linked Data and identified several issues regarding publishing, processing and linking with LOD and non-LOD resources commonly known as “Billion Triple Challenge” as well as the efficient use of semantics in integrating and designing database management systems (Emani et al., 2015). (Nentwig, Hartung, Ngonga Ngomo, & Rahm, 2017) compare multiple link discovery frameworks based on effectiveness (precision and accuracy), efficiency (speed and scalability), and a set of evaluation criteria including linking specifications,

Figure 1. A typical LOD environment
Privacy Implications and Protection in the New Ubiquitous Web Environment
Charalampos Z. Patrikakis, Ioannis G. Nikolakopoulos and Athanasios Voulodimos
[www.igi-global.com/chapter/privacy-implications-protection-new-ubiquitous/39210?camid=4v1a](www.igi-global.com/chapter/privacy-implications-protection-new-ubiquitous/39210?camid=4v1a)

SPedia: A Central Hub for the Linked Open Data of Scientific Publications
[www.igi-global.com/article/spedia/172426?camid=4v1a](www.igi-global.com/article/spedia/172426?camid=4v1a)