Semantic Extension of Query for the Linked Data

Pu Li, Software Engineering College, Zhengzhou University of Light Industry, Zhengzhou, China & School of Computer Science, South China Normal University, Guangzhou, China
Yuncheng Jiang, School of Computer Science, South China Normal University, Guangzhou, China
Ju Wang, College of Computer Science and Information Engineering, Guangxi Normal University, Guilin, China
Zhilei Yin, School of Computer Science, Northwestern Polytechnical University, Xi’an, China

ABSTRACT

With the advent of Big Data Era, users prefer to get knowledge rather than pages from Web. Linked Data, a new form of knowledge representation and publishing described by RDF, can provide a more precise and comprehensible semantic structure to satisfy the aforementioned requirement. Further, the SPARQL query language for RDF is the foundation of many current researches about Linked Data querying. However, these SPARQL-based methods cannot fully express the semantics of the query, so they cannot unleash the potential of Linked Data. To fill this gap, this paper designs a new querying method which extends the SPARQL pattern. Firstly, the authors present some new semantic properties for predicates in RDF triples and design a Semantic Matrix for Predicates (SMᵢ). They then establish a well-defined framework for the notion of Semantically-Extended Query Model for the Linked Data (SEQMᵢ). Moreover, the authors propose some novel algorithms for executing queries by integrating semantic extension into SPARQL pattern. Lastly, experimental results show that the authors’ proposal has a good generality and performs better than some of the most representative similarity search methods.

KEYWORDS

Big Data, Linked Data, Query Model, Semantic Extension, SPARQL

1. INTRODUCTION

With the increasing amount of information, how can we find meaning in these terabytes (Frankel & Reid, 2008)? To answer this question, we should consider from two aspects: On one hand, the World Wide Web (WWW) must be more structured and machine-readable contrasting with the traditional Web. Driven by these demands, Semantic Web (SW) (Berners-Lee et al., 2001) is proposed and widely used, which aims to develop techniques to incorporate semantics into Web design. On the other hand, confronting the sea of online information, users prefer to get knowledge which is more clear and meaningful rather than pages with unstructured text. To counter this requirement, there is a need for new querying techniques to improve traditional keyword-based search (T. Tran et al., 2011).

From the WWW Consortium’s vision of the Web of Linked Data (Hogan et al., 2012), SW presents a revolutionary opportunity for deriving value from data and activity has gained momentum.
with the widespread publishing of structured data as RDF (Klyne & Carroll, 2014). In recent years, an increasing number of data providers like (Bollacker et al., 2007; Hoffart et al., 2013; Lehmann et al., 2015; Nebot & Berlanga, 2016; Wiemann & Bernard, 2016) have published and connected their data into Web of Linked Data and, ultimately, into the SW. Promoted by the eager demand, many Linked Data-oriented techniques have been researched such as (Assaf et al., 2016; Auer et al., 2014; Nguyen & Ichise, 2017; Sande et al., 2016).

The aforementioned developments bring the upsurge to the query for Linked Data. So, the works presented in this paper will make a further study on Linked Data querying from semantics. The rest of the paper is organized as follows. Section 2 briefly reviews related works and highlights the difference between our study and these exiting works. Section 3 introduces some preliminaries about Linked Data querying and provides our research questions and research methodology. In Section 4 we design a well-defined framework as the formalized expression of our query model. The details of processing algorithms of our querying method are presented in Section 5. Section 6 is devoted to evaluating our method. Finally, we draw our conclusion and outline the future work in Section 7.

2. RELATED WORK

The goal of Linked Data query processing is an online execution of declarative queries over the SW, by relying only on the Linked Data principles (Bizer et al., 2009). Aiming to unleash the potential of SW, a number of general methods for querying Linked Data have been developed.

Early querying methods like Sindice (Oren et al., 2008) and Falcons (Cheng & Qu, 2009) are based on keywords. Since SPARQL (Harris & Seaborne, 2010) can be used to express queries across diverse data sources, it becomes the foundation of many recent researches about Linked Data querying. The devices in (Umbrich et al., 2011; Wagner et al., 2012) use index-based source selection and provide source ranking, while the techniques mentioned in (Hartig, 2013; Miranker et al., 2012) are traversal-based query execution approaches without source ranking. Moreover, some domain-oriented or datasource-oriented methods are presented to further pursue the accuracy and efficiency. NAGA (Mahdisoltani et al., 2014) provides best-effort heuristics to return and rank the relevant RDF triples from YAGO (Hoffart et al., 2013). By using SPARQL and SPIN, Lo Bue and Machi (Lo Bue & Machi, 2015) study on integrating and querying over tourism domain datasets via interlinking techniques. MEQLD (Tran & Nguyen, 2015) investigates to improve the mapping extension of lexical entities into DBpedia’s components for creating query in SPARQL. Besides, other current works focus on mapping visual method to SPARQL (Haag et al., 2015), optimizing query results over several heterogeneous Linked Data sources (Taelman, 2016), as well as automatically generating SPARQL query (Alec et al., 2016).

To give a brief overview about the main contributions of the above related works and our studies, the main capacities and features of these similar systems are listed in Table 1.

3. OUR RESEARCH QUESTIONS AND METHODOLOGY

In this section, we will analyze the limitations of the basic SPARQL pattern and present our research questions and the corresponding methodology. Previously, we must introduce some basic concepts for easier understanding subsequent discussions. See especially (Auer et al., 2014; Bizer et al., 2009; Harris & Seaborne, 2010; Klyne & Carroll, 2014; Suchanek et al., 2008) for further details.

3.1. Preliminaries

3.1.1. RDF and SPARQL

RDF defines a data format for representing information in the Web and can be considered as a general model for graph-structured data encoded as triples in the form of subject-predicate-object (s, p, o).
Semantics for the Semantic Web: The Implicit, the Formal and the Powerful
www.igi-global.com/article/semantics-semantic-web/2802?camid=4v1a

Ubiquitous Semantic Applications: A Systematic Literature Review
www.igi-global.com/article/ubiquitous-semantic-applications/113714?camid=4v1a