

Querying Trust in RDF Data with tSPARQL

Olaf Hartig

Humboldt-Universität zu Berlin
hartig@informatik.hu-berlin.de

Abstract. Today a large amount of RDF data is published on the Web. However, the openness of the Web and the ease to combine RDF data from different sources creates new challenges. The Web of data is missing a uniform way to assess and to query the trustworthiness of information. In this paper we present tSPARQL, a trust-aware extension to SPARQL. Two additional keywords enable users to describe trust requirements and to query the trustworthiness of RDF data. Hence, tSPARQL allows adding trust to RDF-based applications in an easy manner. As the foundation we propose a trust model that associates RDF statements with trust values and we extend the SPARQL semantics to access these trust values in tSPARQL. Furthermore, we discuss opportunities to optimize the execution of tSPARQL queries.

1 Introduction

During recent years a large amount of data described by RDF has been published on the Web; large datasets are interlinked; new applications emerge which utilize this data in novel and innovative ways. However, the openness of the Web and the ease to combine RDF data from different sources creates new challenges for applications. Unreliable data could dominate results of queries, taint inferred data, affect knowledge bases, and have negative or misleading impact on software agents. Hence, questions of reliability and trustworthiness must be addressed. While several approaches consider trustworthiness of potential sources of data (e.g. [1,2,3]), little has been done considering the actual data itself.

What is missing for is a uniform way to rate the trustworthiness of the data on the Web and standardized mechanisms to access and to use these ratings. Users as well as software agents have to be able to utilize trust ratings and base their decisions upon them. They have to be enabled to ask queries such as:

- Q₁: Return a list of garages close to a specific location ordered by the trustworthiness of the data.
- Q₂: Return trustworthy reviews for a specific restaurant.
- Q₃: Return the most trustworthy review for each hotel in the city of Heraklion.

To ask queries like these this paper presents appropriate extensions for RDF and its query language SPARQL. Our main contributions are:

- a trust model for RDF data which associates triples with trust values, and
- a trust-aware query language, tSPARQL, which extends SPARQL to describe trust requirements and to access the trustworthiness of query solutions.