Leveraging Terminological Structure for Object Reconciliation

Jan Noessner, Mathias Niepert,
Christian Meilicke, and Heiner Stuckenschmidt

KR & KM Research Group
University of Mannheim, B6 26, 68159 Mannheim, Germany
heiner@informatik.uni-mannheim.de

Abstract. It has been argued that linked open data is the major benefit of the use of semantic technologies on the web as it provides a huge amount of structured data that can be accessed in a more effective way than web pages. While linked open data avoids many problems connected with the use of expressive ontologies, e.g. the knowledge acquisition bottleneck, data heterogeneity remains a challenging problem. In particular, the same objects may be referred to using different URIs in different data sets. Identifying such representations of the same object is called object reconciliation. In this paper, we propose a novel object reconciliation method that is based on an existing semantic similarity measure for linked data. We adapt the measure to the object reconciliation problem, present complete and approximate algorithms that efficiently implement the methods, and present a systematic evaluation of the approach based on a benchmark dataset. As our main result, we show that the use of light-weight ontologies and schema information significantly improves object reconciliation in the context of linked open data.

1 Introduction

There is an ongoing debate concerning the role of ontologies for the semantic web. While rich ontologies have been promoted as an integral part of every semantic web application [10], it is increasingly argued that the real value of the semantic web is based on its ability to create and maintain linked open data which provides effective access to semantically enhanced information on the web [20]. In this paper, we argue that the use of (light-weight) ontologies helps to solve one of the key problems of linked open data on the web, namely, the actual linking of data by identifying different representations of the same object. This problem has been extensively studied in the context of database systems as duplicate detection, record linkage, and object or reference reconciliation [12]. While most existing work has focused on the design of specialized measures which estimate the similarity of objects based on their lexical properties, the use of schema information in the context of formal ontologies has only recently been proposed for instance in [14]. In this work, we leverage schema information to exclude logically inconsistent links between objects and to improve the overall accuracy