

A multi-tiered system for querying the web of data

Ahmed Rabhi, Salah Ouederrou, Rachida Fissoune and Hassan Badir

National School of Applied Sciences, Abdelmalek Essaâdi University, Tangier, Morocco

rabhi.ahmed.1992@gmail.com

salah.ouederrou@gmail.com

ensat.fissoune@gmail.com

hbadir@gmail.com

Abstract. The web of data is a large collection of various data from different domains, these data are distributed on several sources which are heterogeneous and managed independently. Hence, a user may need to query multiple data sources to aggregate pieces of information. In this context, our work aims to set up a system that integrate fragments of information from distributed and independent data sets by providing a single interface to the user. In this paper, we propose a system designed for querying the web of data without having a prior knowledge of the sources contributing to the response. This system adopts a solution for selecting relevant sources in order to decrease network traffic so that the system becomes less dependent on the quality of the connection flow.

1 Introduction

Thanks to Linked Data technologies, the world wide web may enable direct access to raw data, actually Linked Data provides a set of design principles and paradigms for sharing data on the web, thus making the web a global space hosting data in machine-readable format based on standards like RDF and SPARQL. Hence, enabling people and machines to interchange data, which greatly enhance the search for information on the web.

Web of data is a large collection of data from different domains related to each other forming data graphs. According to Heath and Bizer (2011), the web of data is a global space where individuals and organizations have adopted Linked Data standards to publish their data. therefore, the web of data forms a giant graph consisting of billions of RDF data distributed on a large number of Datasets covering all domains such as geography, politics, life science, social networks and others domains. The web of data can be seen as a generic space for sharing containing a variety of data where the user can retrieve information from various domains. These data are stored in RDF datasets and are accessed through sources called Endpoints, the web of data contains nowadays a large number of Datasets like: DBpedia that contain a variety of domains and Bio2RDF that provides a network of linked data for life sciences.

The worry is that the number of data sources has recently increased on a large scale. Moreover, these sources are distributed, heterogeneous and independent one another. Thus, searching information on the web of data means searching information on a number of data sources having no relationship between them. Hence, a user query may require interrogating multiple