KEYSTONE - Based Search on Structured Data Sources

Body:

Motivation

While the Web of data, and in particular linked data, has seen tremendous growth over the past years, the take-up, usage and reuse of data is still limited and is often focused on well-known reference datasets. A lack of scalable and usable methods for formulating and distributing queries across the Web of data poses a challenge and prevents users from obtaining relevant, correct and up-to-date information. The main objective of the KEYSTONE COST Action is to launch and establish a cooperative network of researchers, practitioners, and application domain specialists working in fields related to semantic data management, the Semantic Web, information retrieval, artificial intelligence, machine learning and natural language processing, that coordinates collaboration among them to enable research activities and technology transfer in the area of keyword-based searches of structured data sources. This coordination effort promotes the development of a new revolutionary paradigm that provides users with keyword-based search capabilities for structured data sources such as is currently done with documents.

Challenges & Highlights

Given the dynamic and evolving nature of structured data on the Web, the research community should focus specifically on the development of scalable semantic search techniques for large-scale, distributed and heterogeneous datasets on the Web. These approaches need to be closely aligned with dataset profiling and quality assessment techniques enabling users to select suitable data for a given purpose. The challenges to be addressed include novel scientific methods and techniques for querying, assessing, profiling, and curating distributed datasets as well as the application of perspectives, such as the innovative use of tools and methods for providing structured knowledge about distributed datasets, their evolution, and, fundamentally, means to search and query the Web of data. In this context, we have developed novel scalable techniques to efficiently generate topic profiles for structured datasets and enhance dataset interlinking. Furthermore, we have developed novel applications using open datasets for the enrichment of Web archives. Moreover, we have initiated and co-organized the 1st International Workshop on Dataset Profiling & Federated Search for Linked Data (PROFILES14) during the 11th European Semantic Web Conference (ESWC 2014). The PROFILES14 covered a range of topics related to data source contextualization for search and exploration in Linked Data, profiling of linked datasets, as well as measuring and modelling dynamics and the evolution of Linked Data.

Potential Applications & Future Issues

Application domains that can profit from novel search paradigms for structured data on the Web are numerous and include product search, Web archive enrichment, recommender systems using linked data and many others. In these applications, data consumers are often interested in accumulating entity-centric data spread across distributed structured sources on the Web to obtain consistent and possibly complete entity representations. In the coming year, KEYSTONE will address these issues and further intensify collaborations within its network by means of establishing an open conference, organizing a training school and a number of short term scientific projects. Furthermore, we plan to establish PROFILES workshop as a regular event to raise awareness of KEYSTONE topics within the Semantic Web community. Moreover, a special issue on Dataset Profiling and Federated Search for Linked Data in the International Journal On Semantic Web and Information Systems is in planning.

Project abstract:

The scientific objective of KEYSTONE is to analyze, design, develop and evaluate techniques to enable the keyword-based searching of large amounts of structured data published on the Web. This already amounts to billions of instances distributed across hundreds of datasets and it is constantly growing. This data comes from various sources and domains and has the potential to foster the creation of new services and businesses for political, social and commercial activities. KEYSTONE takes an important step towards cracking the accessibility barrier to the Web of data for web users by bringing together international experts from 27 European countries and developing semantic & keyword-based search techniques for this data.
Members: nejdl
**Project manager:** Dr. Elena Demidova
**Project duration:** Oct 2013 – Oct 2017
**Project research areas:**
- Web Search
- Web Information
- Management

**Project type:** ICT COST Action (IC1302)
**URL:** [http://www.keystone-cost.eu](http://www.keystone-cost.eu)
**Research Area:** Intelligent Access to Information

**Status of the Project:**
- Bibsonomy show project publications: 1
- Bibsonomy use tabs to list publications: 1