QuerioCity: Accessing the Information of a City (DEMO)

Spyros Kotoulas and Vanessa Lopez and Marco Luca Sbодio and Martin Stephenson and Raymond Lloyd and Aris Gkoulalas-Divanis and Pól Mac Aonghusa

Smart Cities Technology Center
IBM Research
Damastown Industrial Estate
Dublin 15, Ireland

Introduction

QuerioCity aims at creating an ecosystem for managing and accessing the information of a city, with a particular focus on transforming, integrating and querying heterogeneous semi-structured data in an open environment.

This raises unique challenges in terms of:

- **Fitness-for-use.** The users of the system are not data integration experts and not qualified to use industry data integration tools. Furthermore, they are not able to query data using structured query languages.

- **Domain modeling.** The domain of the information is very broad and open. As such, generating and mapping data to a single model is infeasible or too expensive.

- **Global integration.** Addressing the information needs for solving problems in an urban environment requires integration with an open set of external datasets. Furthermore, it is desirable that city data becomes easily consumable by other parties.

- **Scale.** The data in a city changes often (streams), is potentially very large and it is interlinked with an open set of external data.

Other approaches in this domain have several limitations: content portals (e.g. twc open gov data portal, data.gov.uk, IOGDC (Ding et al. 2011)) lack sufficient capabilities to explore, navigate and query across collections of multi-domain data; enterprise data integration platforms require significant technical expertise and effort from the user; technical tools for data cleaning and integration (e.g. in (Gonzalez et al. 2010), (Huynh and Mazzocchi )) require some technical skills and lack semantic depth to be able to answer complicated queries.

QuerioCity uses and augments technologies from the fields of Linked Data and Semantics-based Integration. We are tapping on research output from the fields of Pay-as-you-go data integration (based on non-expert input) (Bizer et al. 2009), (Madhavan et al. 2007), RDF stores, Information provenance and Anonymization.

Copyright © 2012, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
“lift” the data as much as possible, as described in the previous section. On the other side, data consumers pull the data in order to fulfill their needs.

There are some issues that span all aspects of the system. Firstly, given the large volume of the data, the number of datasets and the potential number of users, the system has to provide good scalability characteristics. Second, given the open nature of the integration process, it is imperative that the system records information provenance. Finally, given the potentially sensitive nature of the information, the system provides functionality for detection of privacy threats (Fung et al. 2010) and tools for anonymization.

**Demo**

We will perform a demonstration of a prototype version of the QuerioCity platform, as used in the upcoming update of the dublinked.ie website.

QuerioCity demo is based on static and dynamic data spanning a broad range of domains, including transportation, communication, housing, utilities, etc. Such data is provided by Dublin city.

The platform aims at taking raw city data in various formats from data publishers and making them readily accessible as linked open data. In order to do that the platform assists publishers in lifting their data and semantically annotating it. The assisted annotation allows for the creation of meaningful metadata by reusing existent vocabularies, such as dublin core, public sector ontologies and DBpedia (Bizer et al. 2009). The metadata can help us in identifying the type of the content. For instance the content of a dataset annotated with the keyword "parking" can be further refined with the annotations "car park", "car parking permits", "resident parking", "disabled parking", "parking fines", "parking meters", etc. Good metadata allow users, who may not be an expert in a specific domain, to easily explore and mash up data. The platform leverages semantic data types (location, dates, unit of measure) to achieve a uniform type format by automatic conversions. Data can often be contextualized by annotating it with existing vocabularies and by linking it to semantically related entities from other existing datasets (e.g., two datasets about lighting poles in Dublin and Fingal).

On the data consumer side, beyond providing a SPARQL end-point and RESTful APIs to search on the metadata and associated data files, QuerioCity aims at providing a foundation platform for building complex visualisation and analytics explorations of public city data. The platform supports multiple interaction paradigms helping users in navigating information, reducing the search space, exploring data and enabling natural queries that scale. The system helps to discover content and fuse information sources, showing the intersections and relations between previously isolated datasets. Different types of data can be interpreted through different visualisations (Dadzie and Rowe 2011), such as timelines, charts, etc. As city data is situated on a specific temporal and geographical context, further insight is given by comparing datasets through spatial-temporal visualizations or heat maps (optionally points of interest can be extracted from external sources such as Linked open maps).

As such, with QuerioCity the Web of data is brought to its full potential by enriching city data with external data sources, discovering relevant connections across heterogeneous data sources and domains, and understanding the interaction of human behavior with the system and the data (Gonzalez et al. 2010).

**References**


---

1 http://www.dublinked.ie
2 http://dublincore.org/
3 http://doc.esd.org.uk/IPSV/2.00.html
4 http://linkedgeodata.org/About