ICONVIS: an Interactive and Customizable System for Semantic Data Visualization

Federico Cairo
corso Duca degli Abruzzi, 24, Torino
Politecnico di Torino
PhD student
Tel.(+39) 011 0903255
federico.cairo@polito.it

Giuseppe Futia
via Pier Carlo Boggio, 65/a, Torino
NEXA Center for Internet & Society – Politecnico di Torino
Graduate student
Tel.(+39) 011 0907217
giuseppe.futia@polito.it

ABSTRACT

This paper presents an open source software framework that leverages Semantic Web technologies to enrich a data source with new conceptual elements and to visualize contents in a graphical ontology-based fashion. From unstructured datasets such as a database or CSV files, ICONVIS (Interactive and Customizable ONtology-based VIsualization System) allows to semi-automatically build an ontology and display it in the form of a semantic graph. Furthermore, the application enhances database information runtime, retrieving new data from the Linked Open Data (LOD) community through SPARQL queries on the REST endpoints of DBpedia, GeoNames, etc.

Categories and Subject Descriptors

I.2.10 [**Vision and Scene Understanding]** Representations, data structures, and transforms; I.2.4 [**Knowledge Representation Formalisms and Methods**]; H.3.5 **[Online Information Services]** Web-based services.

Keywords

Data visualization, ontology-based data access, ontology browsing, Linked Open Data, information enhancement.

1. INTRODUCTION: AIMS AND BACKGROUND

Among the Italian public administrations, Regione Piemonte has been one of the first to understand the importance of publishing data in open formats. Dati.piemonte.it, the most prominent Italian experience of public open data, aims at facilitating information sharing across the public administration bodies in Piedmont, enabling private citizens to download free datasets in CSV format about a number of public issues, like health, economic activities, education, road conditions, etc.

Nevertheless, these data are open but not linked yet, so different concepts are fragmented and inconsistent to each other, not giving users an overall idea of the wealth of information assets. ICONVIS project aims to bridge the gap between these simple open data and the Linked Open Data [1]. In accordance with the general principles of Linked Data and as a complement to the well-known Tim Berners-Lee's mantra "raw data now", our research proposes data publishers a gradual and at the same time rapid way to enrich their data with semantics and interconnect them on the Web.

ICONVIS is particularly suitable for content publishers, like governments or public bodies, who want to offer their citizens a different and educational utilization of their large data repository. It is a framework that allows to display and browse structured data in the form of an ontology [2]. Ontology is designed on the basis of the most significant values of a data source, so it is a representation of the overall knowledge of the data domain. Users can switch from the exploration of concepts in the graph to the effective fruition of contents of the data source. In addition, ICONVIS is able to reach open datasets on the Web (like DBpedia, GeoNames, Data.gov, etc.) in order to increase information at runtime and to display it to users in a suitable way.

2. DESCRIPTION

The configurability of ICONVIS is due to the presence of XML files where users can specify different parameters. The first is a properties file which is read during the build and which contains information about input ontology, language for the labels of the nodes and a flag for managing DB and LOD modules (activation / deactivation).

```
#Path of "input_files" folder. Example:  C:/someFolder/input_files
ontology.folder.path=

#Ontology file complete name. Example:  someFile.owl
ontology.file.name=test_ontology.owl

# choose a xml:lang value (en, fr, it, ...)
lang=en

#true to activate DB and LOD modules, false otherwise
database.lod.module.active=true
```

An important aspect on which ICONVIS focuses is the mapping, at a query level, between DB contents and ontology nodes. A click on each node of the ontology graph (a class or an instance) produce a DB query to retrieve data of interest to user. Rather than enter information about the DB queries directly in the OWL artifact, we preferred to introduce a configuration file called *DB_query_mapping.xml*. This XML file allows the admin to set the same default behavior in response to a click on every ontology node or to set a custom behavior on a single class or instance. A sample configuration for the *DB_query_mapping.xml* is the following:

```
<?xml version="1.0"?>
<DBmapping>
<class id="owl:Thing">
<class_query>SELECT ... FROM ... WHERE x="#default#"</class_query>
</class>
<indiv id="ex:William_Shakespeare">
<ind_query>SELECT ... FROM ... WHERE x="English poet"</ind_query>
</indiv>
</DBmapping>
```

ICONVIS assigns a default behavior to a click on instances of all ontology classes, except the instance with URI *ex:William_Shakespeare*. The default behavior can be chosen at will by admin: in the ICONVIS basic configuration, software replaces the string *#default#* with the value of the first *rdfs:label* of the ontology node for a chosen language. In this way, the matching between DB and ontology is essentially language-based, since it is based on the correspondence between ontology labels and DB field values. For the individual *ex:William_Shakespeare* software will run a query with the admin specified WHERE clause "English poet".

In the case of retrieving data from LOD repositories, ICONVIS information enhancement module is designed to allow the admin to specify, for a class or each individual, how software should behave as a result of clicking on the RDF shaped icon. This behavior is specified in another configuration file, the *LOD_query_mapping.xml*, as in the example below:

```
<?xml version="1.0"?>
<LODmapping>
<class id="owl:Thing">
<class_endpoint>DBpedia</class_endpoint>
<class_query>SELECT * WHERE {?x rdfs:label "#default#"@it}</class_query>
</class
```

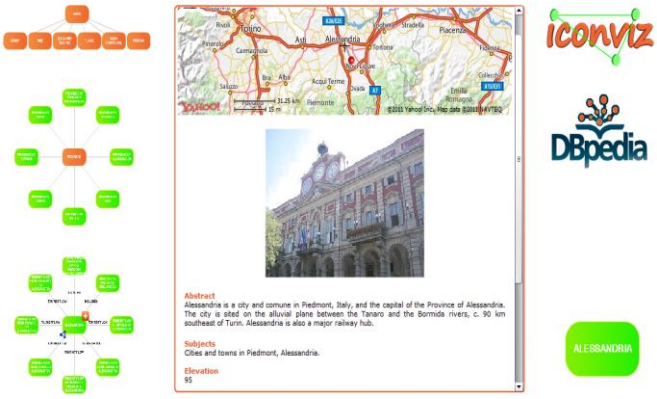
3. APPLICATION AND RESULTS

To realize the ICONVIS demo which is available on our website (http://iconvis.polito.it/iconvis/?q=node/6), we downloaded several CSV dataset from the Dati.Piemonte.it portal and we implemented a domain ontology, formally representing the knowledge contained in these files. Then we automatically populated the knowledge base using the ICONVIS RDF conversion module. Since data from Dati.piemonte.it are not connected to any database, we deactivated database management module and used only the LOD one (DBpedia and GeoNames).

We customized *LOD_query_mapping.xml* in order to "hook" the DBpedia knowledge base with appropriate SPARQL queries. The end result makes the idea of how many advantages the use of ontology offers in the publication of open data. The disconnected and redundant datasets released into the portal have become, through ICONVIS, a consistent and ordered semantic universe. The dizzying CSV lists have been transformed into a series of graphics easy to navigate.

ICONVIS shows all ontology conceptual contents, hiding to user ontology complexity and trying to make the graph simple and intuitive to browse [3]. Instances of each class are displayed with different graphical elements in a radial graph, starting from the navigation of taxonomy. By clicking on a single instance ICONVIS shows in a circle all relations that bind this entity to other entities in the knowledge base. The interface displays labels of ontology classes, individuals and relations in the default language chosen by the admin: when nodes are not labeled, URIs are directly displayed. A mechanism similar to breadcrumbs keeps track of user's navigation path to allow him to return to items previously browsed.

Using ICONVIS demo, users will immediately realize the wealth of information derived from the ontology. Once the user chose the goal of his interest, for example a single school or a single market, a click on the RDF shaped icon procures additional data about that entity taken from DBpedia and GeoNames. New information is displayed in the same navigation interface: in the case of geographic data, Yahoo Maps are exploited to display the area in which the entity is located.



4. CONCLUDING REMARKS

Digitization process has led to the proliferation of a large amount of information, which is more and more complex to manage or exploit. In this context ontologies are a useful mean to restore order, because they describe a collection of concepts related to each other within a specific domain of knowledge. In addition to providing valuable support to semantic search algorithms, ontologies can be represented visually [4]: they become the visual tool through which users can pinpoint information they are interested in, browsing concepts linked by different relationships.

By linking information contained in a data source to information made available on the Web by Linked Open Data providers, ICONVIS actually broadens a content repository beyond its initial capacity and makes it more interesting and rich for users. Our application demonstrates how the use of Semantic Web technologies can help not only to make data interoperable and consistent, but also to find more intuitive and interactive graphical forms for their use.

5. REFERENCES

[1] Heath, T., and Bizer, C. 2011. *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, Morgan & Claypool, pp. 1-136.

[2] Hoekstra, R.J. 2009. *Ontology Representation: design patterns and ontologies that make sense*. IOS Press, Amsterdam.

[3] Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., and Giannopoulou, E. 2007. *Ontology visualization methods a survey*. ACM Computing Surveys. 39, 4, 1–43.

[4] Lanzenberger, M., Sampson, J., and Rester, M. 2010. *Ontology Visualization: Tools and Techniques for Visual Representation of Semi-Structured Meta-Data*, Journal of Universal Computer Science, 16(7), pp. 1036-1054.