Classification Explorer: Navigational Querying of Statistical Classifications

Taeke Gjaltema (UNECE) and Monica Scannapieco (Istat) (editors)[i]

Abstract. In this paper, we describe an application that, starting from statistical classifications represented as Linked Data artefacts, allows for a rich user experience based on cross-classifications navigation.

1   Introduction

Linked Data and Semantic Web standards are being progressively adopted by statistical organizations. Classifications are an indispensable instrument to disseminate statistical data: without classifications no statistics. An integrated system of activity and product classifications allows the comparability of statistics produced in different statistical domains and countries.

The High-Level Group for the Modernisation of Official Statistics (HLG-MOS)[ii] established a project[iii] in 2016 with the use of Linked Data for modeling statistical data as one of its principal aims.

Within this project we developed an application called the Classification Explorer. Starting from statistical classifications represented as Linked data artefacts, it allows for a rich user experience by enabling cross-classification navigation. Several of these classifications were made available for the SemStats challenge. This paper describes how the Classification Explorer application accesses and navigates these classifications.

2   Classification Explorer

The Classification Explorer is an application that provides users with a visual experience to interact with the classifications that were made available by the SemStats challenge (http://semstats.org/2016/challenge/challenge-data). We first describe the main functionalities of the Classification Explorer (Section 2.1) and later the detail of the technological setting of the application (Section 2.1).

2.1   Main Functionalities

The main functionalities of the Classification Explorer are detailed and shown below.

Browsing classifications.

For some of the uploaded classifications (so far 14), a browsing functionality is available that permits to:

- Presenting the overall list of available classifications (see Figure 1).

- Selecting one classification and displaying the details at all the levels (see Figure 2 for different levels' details).

Cross-navigating classifications.

For some classifications, correspondence tables are available, thus permitting to pass from one classification to another. In particular, as shown in Figure 3, by selecting a specific activity of NACEiv classification, through the correspondence table with CPAv, it is it is possible to access all the products in CPA associated to the selected activity. The correspondence tables are defined at all levels of the classifications.

Search among classifications.

The Classification Explorer also has a search function. Given a code of a classification item or a text to search, the functionality returns all the matching classification items for all the classification that are in the database. The results are grouped by each classification they belong to (see Figure 4).

Classifications

- Versione 2007 della classificazione Ateco
- Central Product Classification, Ver.1.1
- Central Product Classification, Ver.2
- Central Product Classification, Ver.2.1
- Classification des produits française - CPF rév. 2.1
- International Standard Industrial Classification of All Economic Activities, Rev.3.1
- International Standard Industrial Classification of All Economic Activities, Rev.4
- Nomenclature d'activités française - NAF rév. 2
- North American Industry Classification System (NAICS) 2012
- Dutch Standaard Bedrijfsindeling (SBI) 2008
- Statistical Classification of Products by Activity, Version 2.1
- Statistical Classification of Products by Activity in the European Economic Community, 2008 version
- Statistical Classification of Economic Activities in the European Community, Rev. 1.1 (2002)
- Statistical Classification of Economic Activities in the European Community, Rev. 2 (2008)

Figure 1: Overall list of available classifications

Figure 2: First and second level detail of NACE Rev 2 Classification

Export to CSV.

In addition to the browsing and navigating functionalities illustrated above, the Classification Explorer also has an exporting functionality available. This functionality permits to export and download the displayed list of items in CSV format.

Figure 3: Cross-classification navigation through the correspondence table between CPA and NACE
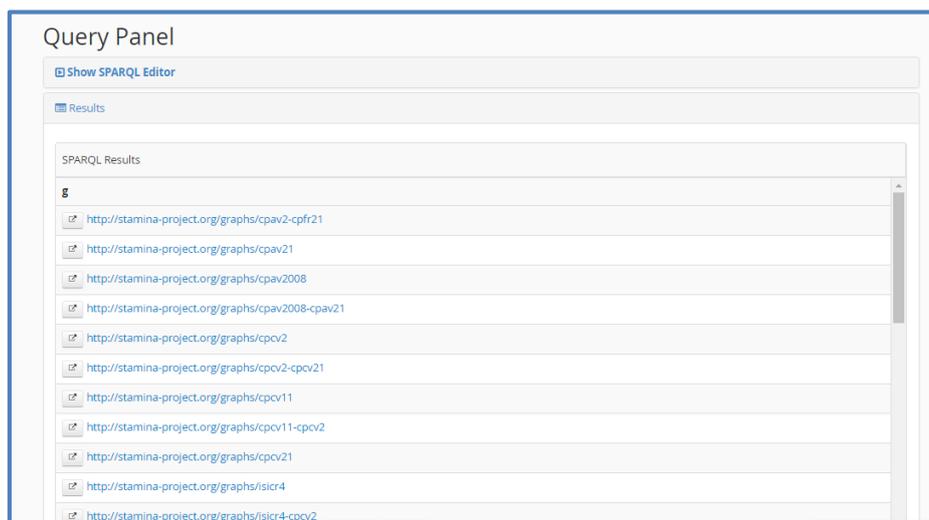

Figure 4: Searching results with searching text "horse"

2.2   Technological Setting

The Classification Explorer is a single-page browser app that communicates with an RDF triple store.

The client is a JavaScript application written using NodeJS (https://nodejs.org/en/), ReactJS (https://github.com/reactjs) for building user interfaces and Webpack (https://github.com/webpack/webpack), a tool to package, deploy and redeploy applications. The client is based on the React-Redux pattern (http://redux.js.org/) that manages the state of the application. As the application is single-page, the navigation is entirely done on the browser. React-router (https://github.com/reactjs/react-router) takes care of updating the URL so that the user can refresh, resume later or share the page she/he is on. It fetches the data from a remote server, detailed in the next section, using SPARQL 1.1 queries over HTTP. Data are then stored locally and so they are accessed only once.

The server-side is an RDF triple store implemented with Stardog (http://stardog.com/). The data made available within the SemStats challenge have been all uploaded to the triple store. These are RDF data, modelled according to several vocabularies, the principal of which is XKOS. In particular, the classifications are organized in a single database with a named graph for each specific classification version. Figure 5 shows the result of a query for retrieving all the graphs.

Figure 5: List of the named graph in Stardog database

3   Conclusions and Final Remarks

The design and implementation of the Classification Explorer was driven by the idea that it is useful to have an integrated access to the various official classifications. Moreover, by exploiting the available correspondence tables, the Classification Explorer has the important advantage of enriching the user experience when accessing classifications by allowing a cross-navigation among them. The application code is available at: https://github.com/UNECE/Classification-Explorer.

When implementing the Classification Explorer, there were some issues related to the lack of harmonized design choices for RDF triples generation. For instance, in the European classifications (NACE and CPA, two versions of each), we had to fix the fact that they used xkos:hasLevels instead of xkos:levels.

The Classification Explorer can be enriched in several ways, namely:

- The layout of the HTML pages can be improved. Indeed, so far only a basic layout is adopted.
- The navigation features can be improved by using URIs for concepts representing items in the correspondence tables instead of strings that prevent from a full navigation.
- Enriching the export and download functionality by having bulk downloads in different formats of the available classifications.

---

[i] The application and the paper were a joint product of the participants of the Sprint Workshop of the UNECE HLG Linked Statistical Metadata project: Raffaella Aracri, Mauro Bruno, Franck Cotton, Eric Debonnel, Taeke Gjaltema, Dennis Grofils, Hans van Hoof, Olivier Levitt, Enrico Orsini, Andrea Pagano, Jean-Baptiste Rudant, Monica Scannapieco, Romain Tailhurat, Laura Tosco and Luca Valentino.

[ii] High-Level Group for the Modernisation of Official Statistics (HLG-MOS) was established in 2010 to oversee and co-ordinate international work relating to statistical modernisation. It promotes standards-based modernisation of official statistics. The aim is to improve the efficiency of statistical production, and the ability to produce outputs that better meet user needs. One of the ways to collaborate and to make progress is through HLG projects in specific areas.

[iii] Implementing ModernStats Standards Project: http://www1.unece.org/stat/platform/display/hlgbas/Implementing+Modernstats+Standards

[iv] The Statistical classification of economic activities in the European Community, abbreviated as NACE, is the classification of economic activities in the European Union (EU); the term NACE is derived from the French *Nomenclature statistique des activités économiques dans la Communauté européenneNomenclature statistique des activités économiques dans la Communauté européenne*

[v] The Statistical classification of products by activity, abbreviated as CPA, is the classification of products (goods as well as services) at the level of the European Union (EU).