

# Analysis of XBRL documents representing financial statements using Semantic Web Technologies

Sheila Méndez Núñez<sup>1</sup>, Jose Emilio Labra Gayo<sup>1</sup>,  
Javier de Andrés Suárez<sup>2</sup>, Patricia Ordóñez de Pablos<sup>2</sup>

<sup>1</sup> Department of Computer Science, University of Oviedo, C/Calvo Sotelo S/N,  
33007 Oviedo, Asturias

<sup>2</sup> Department of Business Administration and Accounting, University of Oviedo, Avda. Del  
Cristo S/N,  
33071 Oviedo, Asturias  
sheilamendeznunez@gmail.com  
{labra, jdandres,patriop}@uniovi.es

**Abstract.** This paper presents an approach to analyze XBRL documents using semantic web technologies. The system takes an XBRL document and converts its information into RDF files. The obtained files are merged with an OWL ontology describing financial information domain. The system enables the formulation of SPARQL queries over the generated data which facilitate the analysis of the financial information. Currently the system has been applied to the XBRL reports generated by the Spanish Securities Commission.

**Keywords:** semantic web, XBRL, financial information, Spanish Securities Commission

## 1 Introduction

XBRL (eXtensible Business Reporting Language) is an XML based standard developed by a not-for-profit international consortium<sup>1</sup> of approximately 450 organizations including regulators, government agencies and software vendors. It has successfully been applied in the exchange and representation of business information between different organizations. XBRL employs XML Schema and XLink technologies to describe different taxonomies for specific domains so that each XBRL document is an instance of an specific XBRL taxonomy.

In Spain, XBRL adoption was led by the Bank of Spain (Banco de España) and by the Spanish Securities Commission<sup>2</sup> in 2005. In the case of the Spanish Securities Commission, it publishes periodically via web, in XBRL format, financial information regarding the listed firms. In our paper we will concentrate in the IPP (Public Periodic Information - *Información Pública Periódica*) taxonomy, which contains the main items of the annual accounts.

---

<sup>1</sup> <http://www.xbrl.org>

<sup>2</sup> <http://www.cnmv.es>

On the other hand, the semantic web can be defined as a long-term vision which pursues the development of technologies that facilitate the automatic manipulation of data published on the Web. Led by the World Wide Web consortium<sup>3</sup>, a number of semantic web technologies have appeared, like RDF to describe resources using a graph model, OWL to define ontologies based on description logics, and SPARQL, to define queries over data RDF graphs. One important aspect of these technologies is that they can be neatly combined using several tools and even allowing the system to infer new knowledge using description logics capabilities.

## 2 Architecture of the system

The architecture of the system is depicted in figure 1.

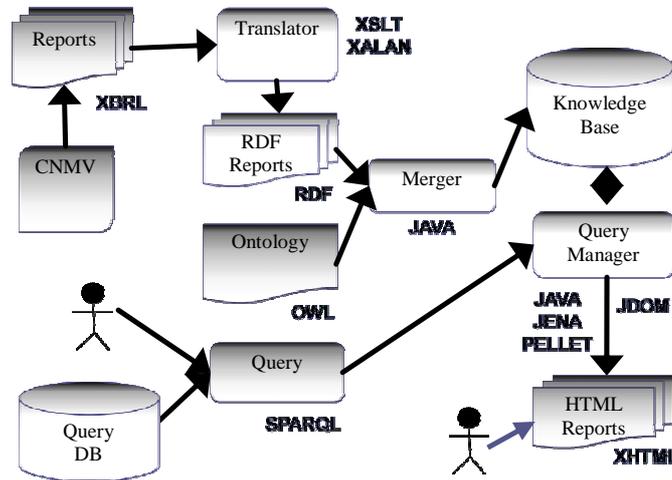


Figure 1: Architecture of the System

The main components are:

- **Translator:** It takes the XBRL reports produced by the Spanish Securities Commission (*Comisión Nacional del Mercado de Valores –CNMV–*) and converts them to RDF reports. We use XSLT to carry on the conversion. Each XBRL concept is transformed into a OWL class or property. Using XSLT improves flexibility and standardization over other alternatives like JDOM, which depend on the Java programming language.
- **Merger:** It combines the RDF reports obtained from the translator with an domain specific OWL ontology. The ontology defines over 100 concepts from the Accounting Principles issued by the International Accounting Standards Board. The results of the merger are stored in the knowledge base of the system.

<sup>3</sup> <http://www.w3c.org>

- **Query Manager:** This module executes SPARQL queries over the knowledge base. It is possible to export the results in HTML reports. The queries can be defined by the end user or they can also be stored and retrieved from a query store.

The system has been implemented as a standalone Java application, although we are planning to develop a web based interface.

### 3 Related Work

The XBRL approach to define taxonomies using XML Schema and XLink has already been criticized by semantic web practitioners, and there have been several proposals to combine both technologies. Recently, the XBRL Ontology specification group has been created to that end [2].

Our approach has been inspired by [1], however, in that paper, the authors develop an ontology over the financial information which is specific of investment funds, while we concentrate in the financial statements of listed firms which can be obtained through the CNMV website.

### 4 Conclusions

In this paper we have presented the architecture of a system which facilitates the analysis of XBRL documents. The system has already been implemented and can be used to obtain several practical results. As an example, it can be used to assess the financial health of an enterprise, to assess its future profitability, or to search enterprises by a given profile.

The proposed system offers a great flexible way to adapt to other domains and we are planning to apply it to other XBRL taxonomies. At the same time, we are also considering to develop a web site which could offer several services for the accounting community.

### References

- 1 R. Lara, I. Cantador, P. Castells: XBRL taxonomies and OWL ontologies for investment funds, First International Workshop on Ontologizing Industrial Standards at the 25th International Conference on Conceptual Modeling (ER2006), Tucson, Arizona, USA, November 6-9, 2006
- 2 XBRL Ontology Specification Group. <http://groups.google.com/group/xbrl-ontology-specification-group>