

Semantic Web Technologies for Economic and Financial Information Management*

J. L. Alonso¹, C. Carranza², P. Castells², B. Foncillas¹, R. Lara³, M. Rico²

¹ Tecnología Información y Finanzas
C/ Españolto 19, 28010 Madrid
{jalonso, bfoncillas}@afi.es

² Universidad Autónoma de Madrid
Ctra. de Colmenar Viejo km. 15, 28049 Madrid
{cesar.carranza, pablo.castells, mariano.rico}@uam.es

³ Universität Innsbruck, Institut für Informatik (IFI)
Techniker Strasse 13, 6020 Innsbruck, Austria
ruben.lara@uibk.ac.at

Abstract

We present an ontology-based platform for economic and financial content management, search and delivery. Our goals include a) the development of an ontology for the domain of economic and financial information, b) the integration of contents and semantics in a knowledge base that provides a conceptual view on low-level contents, c) an adaptive hypermedia-based knowledge visualization and navigation system, and d) semantic search facilities.

1 Introduction

The field of economy and finance is a conceptually rich domain where information is complex, huge in volume, and a highly valuable business product by itself. A massive amount of valuable information is produced worldwide every day, but no one is able to process it all. Efficient filtering, search, and browsing mechanisms are needed by information consumers to access the contents that are most relevant for their business profile, and run through them in an effective way.

The finance community is a major spender in information technology. The web has created new channels for distributing contents, to which more and more activity and information flow has been shifting for more than a decade. The new web technologies are enabling a trend away from monolithic documents, towards the emergence of new content products that consist of flexible combinations of smaller content pieces, fitting different purposes and consumers, and procuring a more efficient capitalization and reuse of produced contents.

Along this line, a number of XML standards for financial contents and business have been defined during the last few years, like FpML, XBRL, RIXML, ebXML, NewsML, IFX, OFX, MarketsML, ISO 15022, swiftML,

MDDL, to name a few [Coates, 2001]. Most of them are concerned with describing business processes and transactions. Some, like XBRL, RIXML and NewsML, do focus on content structure and provide a rich vocabulary of terms for content classification. Our assessment is that these vocabularies need significant extensions when faced to the actual needs of content managers that deal with advanced financial information. More insightful semantics and a sharper level of representation are required to describe and exploit complex information corpora.

The purpose of our work is to achieve an improvement in current Internet-based economic information management practice by adopting Semantic Web technologies and standards in a real setting. We have undertaken a joint project involving a content provider in this field, and two academic institutions, aiming at the development of an ontology-based platform for economic and financial content management, search and delivery. The specific technical goals of this project are:

- Define an ontology for the economic and financial information domain.
- Develop ontology-aware tools for content provision and management.
- Develop a hypermedia-based module for content visualization and semantic navigation in web portals.
- Support semantic search in terms of the economic and financial information ontology.
- Include a user modeling component to be used in navigation and search.

2 Financial and Economic Information Providers

Tecnología, Información y Finanzas (TIF), is part of a company corporation that generates high-quality economic information (equity research notes, newsletters,

* This work is funded by the Spanish Ministry of Science and Technology, grants FIT-150500-2003-309, TIC2002-1948.

analysis, sector reports, recommendations), and provides technologic solutions for information consumers to access, manage, integrate and publish this information in web portals and company intranets.

The consumer profile of this information is diverse, including financial institutions, banks, SMEs that use the information in decision making and foreign trade activity, and distributors who publish the information in first-rank printed and digital media about Spanish economic activity. Adequating the information and delivery procedures to such heterogeneous customer needs, interests, and output channels, is quite a challenge.

A large group of professionals and domain experts in the company is in charge of generating daily economic, market, bank, and financial analyses, commercial fair reports, import/export offers, news, manuals, etc. This information is introduced in the company database, which feeds the automatic delivery systems and web sites. Contents are organized and processed on the basis of a conceptual model (in expert's mind), a vocabulary for information structures and classification terms, which is driven by market needs and reflects the view of the company on the information products it deals with. This model is present somehow in the current TIF software system for information management: it is implicit in the design of the database. As a consequence the possibilities to reason about it are fairly limited.

3 A Semantic Knowledge Base for Economic and Financial Information

Our first endeavor in this project is to wrap the current databases where contents are stored into a knowledge base that provides a conceptual ontology-based view of the information space, above the low level content storage system.

We have built an ontology where the conceptual model of TIF is explicitly represented. It includes concepts like MutualFund, IndustrySector, CommercialFair, EconomicIndicator, CompanyReport, TechnicalAnalysis, FinancialAnalyst, Publisher, Association, and Business-Opportunity, relations between such concepts, and several classification hierarchies for subject topics, industry sectors, intended audience, and other content fields. In this ontology, the old data model has been transformed and augmented with explicit semantics, and enriched with collected domain expertise from TIF. We have integrated the RIXML classification schemes as well, extending and adapting them to support the TIF concepts, terminology, and views. We have defined a mapping from our ontology to RIXML and NewsML formats. The conversion from our ontology to these standards implies a (meta)information loss, in exchange for a wider potential dissemination.

The knowledge base can be queried and browsed directly in terms of the conceptual view. Meaningful queries can be expressed in terms of the vocabulary provided by the ontology, improving current keyword-based search. The database from which actual contents and data

are retrieved has not been redesigned, which would have implied a major cost and a disruption for a critical service that needs to keep going. Instead, we have developed a gateway that dynamically maps ontology instances to (combinations of) database records and fields.

The tools for inputting contents have been adapted to allow defining richer semantics in terms of the ontology. Content managers themselves are users of a highly expressive version of the search and browsing facilities. Efficiency and precision in locating the right contents, and ease of navigation through them, are essential for authors who classify and link pieces together to define global information structures.

The explicit ontology allows more meaningful and precise user profiles, which can express preferences on specific topics, content classes, or even abstract content characterizations. User profiles are taken into account by the adaptive hypermedia-based visualization and navigation module, which is based on our previous work on Pegasus [Castells, 2001]. It uses an explicit presentation model, defined in a fairly simple language, where parts of a semantic network can be easily referenced, and conditions over the user model can be expressed. Presentation models are associated to ontology classes, and define what parts (attributes and relations) of a class instance must be included in its presentation, their visual appearance and layout.

5 Conclusions

The development of a significant corpus of actual Semantic Web applications has been acknowledged as a necessary achievement for the Semantic Web to reach critical mass [Haustein, 2002]. The project presented here intends to be a contribution in this direction. It takes up our previous research work on Semantic Web user interfaces and adaptive navigation systems [Castells, 2001], and will provide a testing ground for our past and future research.

The system is currently under active development. Protégé, RDF(S), Jena 2, and RDQL are used to build, represent, parse, and query the ontology. A full implementation of the system is scheduled to be released by the beginning of 2004.

References

- [Castells, 2001] P. Castells and J. A. Macías. An Adaptive Hypermedia Presentation Modeling System for Custom Knowledge Representations. *World Conference on the WWW and Internet (WebNet 2001)*. Orlando, 2001.
- [Coates, 2001] A. B. Coates. The Role of XML in Finance. *XML Conference & Exposition 2001*. Orlando, Florida, December 2001.
- [Haustein, 2002] S. Haustein and J. Pleumann. Is Participation in the Semantic Web too Difficult? *International Semantic Web Conference (ISWC 2002)*. Sardinia, Italy, 2002.