Semantic Web Technologies for Economic and Financial Information Management

Pablo Castells[1], Borja Foncillas[2], Rubén Lara[3], Mariano Rico[1], Juan Luis Alonso[2]

[1] Universidad Autónoma de Madrid
http://nets.ii.uam.es/
{pablo.castells,mariano.rico}@uam.es
[2] Tecnología, Información y Finanzas
http://www.grupoanalistas.com/
bfoncillas@afi.es
[3] Digital Enterprise Research Institute (DERI) Innsbruck
http://deri.semanticweb.org/
ruben.lara@uibk.ac.at

Abstract. 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. Novel management techniques are required for economic and financial information in order to enable an efficient generation, management and consumption of complex and big information resources. Following this direction, we have developed and ontology-based platform that provides a) the integration of contents and semantics in a knowledge base that provides a conceptual view on low-level contents, b) an adaptive hypermedia-based knowledge visualization and navigation system and c) semantic search facilities. We have developed, as the basis of this platform, an ontology for the domain of economic and financial information.

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 world-wide every day, but its processing is a hard and time-consuming task. 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 the contents produced.

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, among others [4]. Most of them are concerned with describing business processes and transactions. Some, like XBRL [16], RIXML [15] and NewsML [13], 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.

Currently, most of the economic and financial information generated by information providers is mainly textual and, therefore, it cannot be interpreted and processed by computers. This leads to the same problems the management of current Web contents is presenting nowadays. [5] summarizes these problems in the following major points: searches are imprecise, yielding an excessive number of matches; information consumers face the task of going through a big volume of matches in order to get the information required; in addition, the maintenance of the information resources is complex.

The Semantic Web [2] aims at overcoming the problems summarized above by providing an explicit representation of the semantics underlying information sources. Ontologies [8] constitute the backbone technology for the semantic web and, more generally, for the management of formalized knowledge in the context of distributed systems. They provide machine-processable semantics of data and information sources that can be communicated between different agents. Information is made understandable for the computer, thus assisting people to search, extract, interpret and process information.

Semantic Web technologies can naturally be applied to the domain of economic and financial information in order to overcome its current limitations regarding information management. 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 [1]. The specific technical objectives of this project are:

- Define an ontology for the economic and financial information domain that must solve the needs of both the content provider and the information consumers.
- 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 in order to improve the quality of the results.
- Include a user modeling component to be used in navigation and search.
- Easy to adopt solution for the content provider i.e. improve the steps in the current business process but without major changes in the overall process.

This paper presents a real use case in the field of economic and financial information management, its limitations when dealing with current contents, and the approach we have followed to build an ontology-based tool meeting the domain requirements. The paper is structured as follows: section 2 presents our working domain; section 3 details the approach followed and the tools developed so far; section 4 conducts a discussion about our experiences in the project and the main results achieved; finally, section 5 summarizes the conclusions of our work and points out its limitations and future extensions.

2 Description of the domain

Tecnología, Información y Finanzas (TIF)[1] is part of a corporation that generates high-quality economic information (equity research notes, newsletters, analysis, sector reports, recommendations), and provides technology 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 a big challenge.

A wide group of professionals and domain experts in the company is in charge of daily generating a wide range of valuable economic and financial information, including economic, market, bank, and financial analyses, commercial fair reports, import/export offers and news and manuals, among others.

A number of custom web-based content management systems are used for the different types of information generated in the organization. They support the user in creating, editing and publishing this data. The information generated is introduced in the company database, which feeds the automatic delivery systems and web sites.

The custom content management systems are based on web forms that request the appropriate information for a given content type. These forms are created following an ad-hoc procedure, and they are not related to any explicit conceptual model in the company, only to the correspondent part of the database schema.

Contents are organized and processed on the basis of a (mental) conceptual model, 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 provides. This model is present somehow in the current TIF software system for information management, and it is implicit in the design of the database.

Therefore, the semantics of the information stored in the company and its structure is not clearly defined. This makes the generation and maintenance of content managers a time-consuming and error-prone task. Furthermore, the selection of the appropriate information to publish on the web portals and to deliver to the customers is not

[1] http://www.grupoanalistas.com/

a trivial task, as the intended meaning of some of the information in the database is not easy to interpret.

In this context, our assessment is that the procedures followed in the company can be greatly improved by means of the construction of an explicit and formal definition of the conceptual information model of the organization. Such an unambiguous model would provide a uniform view of the information stored in the company, potentially bringing the following benefits: improvement of the data quality, uniform interpretation of the information by providers and consumers, reduction of the information maintenance effort, and semi-automatic generation of new web portals and delivery systems based on the conceptual model and the user profiles.

3 An ontology-based approach to the information management problem

In order to overcome the current limitations in the management, access and search of the information generated by TIF, we have developed an ontology-based platform that applies semantic web technologies in a real setting and shows the usefulness of such technologies. The main components of the platform architecture (see Fig. 1) are: the economic and financial information ontology, the import (from the TIF database) and export (to different formats) facilities, the content management and provision tools, the visualization interface and the search interface and engine. Each of these components is detailed in the following sections, together with the motivations of the approach followed and the choice of technologies.

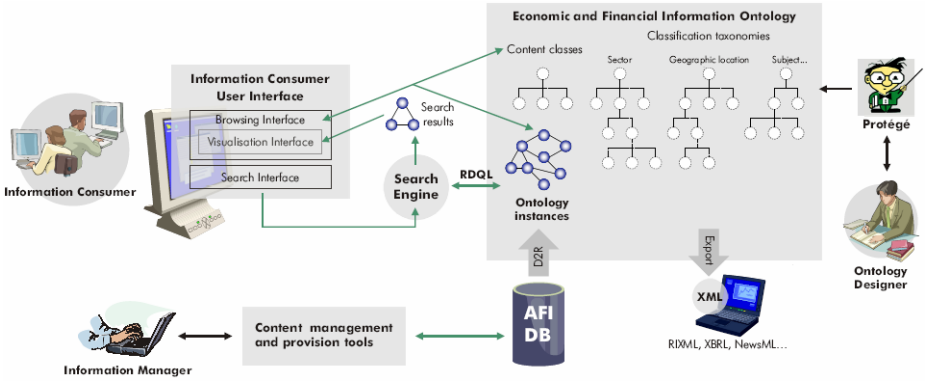

Fig. 1. System architecture

3.1 The economic and financial information ontology

The role of the economic and financial information ontology is central in our architecture, as it reflects the explicit conceptual model followed in the organization for the generation, management and access of the contents provided. The design of the

ontology has been driven by the TIF domain and requirements. The possibility of reusing other ontologies or taxonomies, such as GICS for the classification of sectors, was considered. However, none of them met the TIF requirements, so the reuse of other standards had to be discarded.

The procedure followed to design the domain ontology was incremental, interacting with domain experts in order to produce refined versions of the ontology. Two main steps can be distinguished in the design:

1. First version of the ontology, which was designed based on the existent database schema.
2. Interaction with domain experts from TIF in order to refine the ontology, addition of missing concepts, relations and properties, and assurance of the appropriate coverage of the domain.

The interaction with the domain experts has been the most crucial step for a successful design of the ontology, as they have contributed with numerous and valuable improvements to the first version of the domain ontology. The first step was motivated by the need of making explicit the current structure of the database in order to use it as the starting point for the interaction with the domain experts.

Ontology characteristics

The requirement analysis studies carried out with in-house financial and technical experts have led us to establish four distinct kinds of concepts (classes) in the developed ontology:

1. Content classes. They stand for information products created by financial experts at TIF. Each TIF information product is described by an instance of a content class.
2. Classification categories. No instances are created for these classes. The classes are used directly as values for the *category* property of content instances. The categories form a taxonomy that serves as a classification scheme.
3. Entity classes. They represent all other information items that are not produced by financial experts, but that are used to annotate contents. This includes concepts like companies, banks, organisations, people, information sources, event hosting facilities, etc.
4. Enumerated types for certain property values. They provide sets of values (controlled vocabularies). These classes contain just the *value* and *code* properties, and have a fixed and moderate number of instances.

From our experience in ontology engineering for information systems, the consideration of these four kinds of classes is an interesting and recurrent distinction that arises in many, if not most, information management systems in diverse domains. In fact, although perhaps not explicitly stated, a similar approach can be found in information exchange standards like RIXML [15] and other standards in the controlled vocabulary community [9]. As is usually the case when attempting a subdivision of the knowledge representation primitives, the distinction is not necessarily always a sharp line. Our proposed scheme responds to a careful study of experts' and users' needs and domain understanding, information system development know-how, and the capabilities of the underlying technological support (e.g. web-based navigation, internal information organisation and storage).

The developed ontology provides explicit connections between contents, categories, and other entities, that were only implicit in the current implementation. These relations are now well characterized, and can be further described in as much detail as needed, as the employed semantic web technology allows. As will be described later, this is exploited in our platform to support more expressive and precise search capabilities, and for the semi-automation of the generation of user interfaces and forms for search, information visualisation, and content provision.

Ontology language

Regarding the choice of the ontology language used to define the ontology, the following criteria were followed:

- Maturity of the language, including its degree of standardization.
- Sufficient tool support for the design and maintenance of the ontology.
- Appropriate expressiveness for modelling of the domain.

Following these criteria, RDF(S) was chosen. The rationale behind this choice is:

- RDF(S) is a W3C recommendation, which provides a guarantee of its maturity and stability. OWL was also considered, but as it is still in the process of becoming a W3C recommendation, its stability was not clear enough.
- RDF(S) is the most widely supported language by the available tools. That fact reduces the risks in the development and the time necessary to implement our architecture.
- After numerous meetings with domain experts from TIF, it was shown that the expressivity of RDF/RDFS was enough for the TIF business activity. The transitive closure of subPropertyOf and subClassOf relations, the domain and range entailments, and the implications of subPropertyOf and subClassOf, were the only inference mechanisms used. These are supported by the "simple" RDFS inference level of Jena. No further expressivity or inference mechanisms were required to meet the provider requirements for the generation of the information and the requester requirements for its consumption.

Tool support

Among the ontology development tools available with support for RDF, Protégé-2000[2] was selected because of its maturity, ease-of-use and, what is more important, its scalability and extensibility.

Protégé-2000 has thousands of users all over the world who use the system for projects ranging from modelling cancer-protocol guidelines to modelling nuclear-power stations [6]. It provides a graphical and interactive ontology-design and knowledge-base development environment. It helps knowledge engineers and domain experts to perform knowledge-management tasks. Ontology developers can access relevant information quickly whenever they need it, and can use direct manipulation to navigate and manage an ontology.

In addition to highly usable interface, two other important features distinguish Protégé-2000 from most ontology-editing environments: its scalability and extensibility. Developers have successfully employed Protégé-2000 to build and use ontologies

[2] http://protege.stanford.edu

consisting of 150,000 frames. Supporting knowledge bases with hundreds of thousands of frames involves two components: (1) a database backend to store and query the data and (2) a caching mechanism to enable loading of new frames once the number of frames in memory has exceeded the memory limit.

One of the major advantages of the Protégé-2000 architecture is that the system is constructed in an open, modular fashion. Its component-based architecture enables system builders to add new functionality by creating appropriate plugins. The Protégé Plugin Library[3] contains contributions from developers all over the world. Plugins for other ontology languages such as DAML+OIL and OWL assures an easy evolution of our ontology if a higher expressiveness is required in the future.

The result of our conceptual work in cooperation with the domain experts, using RDF(S) as ontology languages and Protégé-2000 as the ontology development tool, is an ontology that has been approved by the company as appropriately reflecting its business domain and that fits to the huge volume of information already present in the organization i.e. the current information can be easily expressed in terms of the ontology. Fig. 2 partially shows the resulting ontology. All the elements of the ontology are described in Spanish. No multilingualism support has been considered, as the business activities of TIF are mainly focused on the Spanish market.

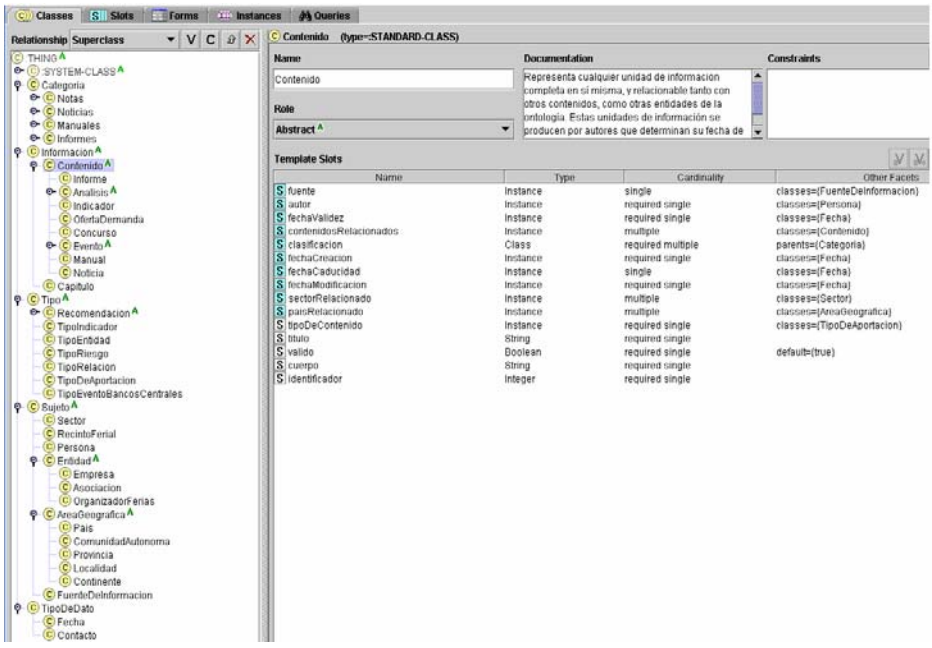

Fig. 2. Ontology in Protégé-2000

[3] http://protege.stanford.edu/plugins.html

3.2 Linking to existing data

As exposed in the previous sections, TIF has a number of experts in economy and finances that daily generate economic and financial information that is stored in the TIF information systems. Several applications make this information arrive to the customers for its consumption, either via the information delivery systems (XML delivery) or via the various web portals TIF feeds with this information. Our project has been conceived as an extension and improvement of the current exploitation system. Therefore, all the information already available in the organization had to be annotated using the ontology described in section 3.1, while keeping some other applications in the company (assessment systems, knowledge management system, etc.) working properly. For this reason, in addition to the annotation of the information for the delivery systems and web portals, the information has to be also stored in the company databases to keep these other systems running.

Our solution was designed to achieve the improvement of our target applications (delivery systems and web portals) with the restriction of not having any impact in other company applications.

The first step was to annotate the contents already available using our domain ontology. For this task, we used the open source tool D2R. D2R connects to RDBMSs supporting JDBC or ODBC and, using an XML mapping file, extracts the information from the RDBMS and generates RDF instances.

An XML mapping file has to be created for each concept in the ontology. The file defines how the results of an SQL query on the RDBMS are mapped to the concept attributes. An ontology concept requires in some cases information from several database entities. How this information is gathered is defined in the SQL query.

We have generated such mappings and, from the available contents, the appropriate RDF instances have been created. These instances have been in turn stored in the organization RDBMSs to ensure the persistency of the data. For this purpose, Jena has been used. Jena retrieves the RDF instances of the ontology from the files generated using D2R and stores them in the RDBMSs.

A problem that we faced was how to, after the annotation of available information, deal with the new information daily generated in the company. As we had the constraint of keeping all the information (existing and to be generated) stored in the existing database schema for its use by other applications, we had to maintain two copies of the organization information. In order to reuse the mappings already defined for the information annotation of existing data, the new information generated is first stored in the RDBMSs and then such mappings are used by D2R to annotate this information. The ontology instances are then stored in the RDBMSs.

3.3 Information search

Our platform provides a search module where customers, content providers, content managers and administrators can query the knowledge base. Our search module improves the facilities provided by the information management system version running at TIF before our project was started in several ways. Whereas the former system only provided keyword-based, full-text search, and a simplified, ad-hoc, partial form

of structured search, our module supports full structured search in terms of any dimension of the ontology, and allows setting different levels of detail and difficulty of use, depending of the intended user profiles.

In our system the user interacts with an HTML search form interface where s/he can select concept types in the ontology (content classes), and provide search keywords for properties of the class. Thus the user can formulate expressive information needs in terms of classes, properties, and relations among contents and concepts. The search forms are automatically generated by a generic mechanism from ontology class descriptions. The search form generation mechanism shares much functionality with (actually it can be viewed as a particular case of) the visualization module, described in the next section. Here we describe the features that are specific to the search forms generation.

The search form generation mechanism provides a default procedure to generate forms adapted to the structure and field types of classes, and the possibility to define custom form design by means of search form templates for classes. For the default procedure, the properties of content classes have a boolean "searchable" metaproperty, with which ontology designers can control whether or not the generated search forms should include an input control where search values for the property can be supplied. The generation procedure selects different HTML/JavaScript controls depending on the type of the searchable property.

The default mechanism provides an instant search facility as a by-product of the ontology and knowledge base construction. However, it is usually necessary to create a custom form design in accordance with the global application look and feel and brand image considerations. This is achieved in our system by creating form templates for each content class, where all aspects of the design can be defined in as much detail as desired. Our template definition language is based on JSP, where custom tags have been defined to provide a simple vocabulary for expressing property references and other ontology graph traversal expressions. The language also includes primitives to easily specify HTML or JavaScript input components, and facilities to define global layout constructs. Wherever details are not explicitly indicated in the template, the system tries to provide appropriate default solutions. An example of the semantic search form for a given class of the ontology is given in Fig. 3.

It has been studied that it is generally more adequate to provide customers with fairly simple and easy to use search interfaces [7], whereas experts and content managers, who are aware of many internal information details, can benefit from more complex and powerful search facilities. This is supported in our platform by creating different templates for different user profiles and usage modes, thus enabling the creation of as large and varied an array of power levels and modalities as needed (see [12] for an overview of user interface approaches for searching), in a highly modular way, very easy to extend. In our current implementation, the set of searchable properties and the search form design for each class have been provided by information management experts and graphic designers from TIF.

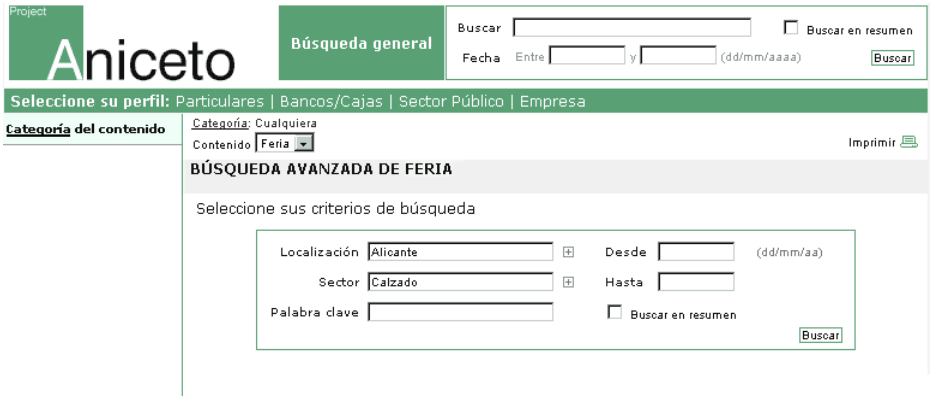

Fig. 3. Semantic search form for a "Feria" concept

The possibilities to use the ontology vocabulary to formulate information needs in our system go beyond specifying property values for content classes. The search module allows the user to combine direct search, using content classes and fields, with navigation through the classification taxonomies included in the ontology. This approach follows the classic combination of searching and browsing in systems like Yahoo! and others [11]. The user can restrict the direct search to selected taxonomy categories. With the search results, the system returns the list of categories to which the results belong, using which the user can narrow or widen his/her search, or go back to browsing.

The search module converts the information need conveyed by the user into an RDQL query, which is executed against the ontology, yielding a set of RDF instances that match the query. The list of instances is presented to the user in an understandable way in a web page. The user can click on instances, which are then displayed in detail in a full page (or a large page area). The way individual instances and lists of search results are presented to the user is controlled by a visualization module that is described next.

3.4 Information visualization

The results of the search are a list of ontology instances that satisfy the information need expressed by the user. Our platform includes a specialized module to present this information and allow the user to run through them, visualize the information units, and navigate across units. This module is based on our early work on the Pegasus tool [3].

The visualization module shows instances of the ontology in dynamically generated web pages. Each class of instance is presented in a different way, showing its data and relations to other instances selectively. Instead of hardwiring this treatment in a program, our platform allows defining the presentation of each ontology class independently, using one or several visualization models per class.

The presentation model of each class establishes the parts of an instance that have to be shown, in what order and under what appearance. This model is defined with a fairly simple language that permits referencing and traversing easily the parts of the semantic network that have to be visualized. The presentation engine selects dynamically the appropriate view for an instance at the time it has to be presented, according to the instance class, and other conditions, if any. The visualization module takes also care of presenting in the same page other instances related with the one being visualized, or of generating hyperlinks to them instead, in order to navigate across ontology relations.

The presentation language is based on JSP, with a library of custom tags which allow creating, besides free HTML and JavaScript code, a) ontology access expressions, b) HTML / JavaScript primitives that display ontology constructs, and c) layout constructs. The presentation models currently defined for the ontology classes have been constructed by inserting the appropriate ontology references and presentation constructs, into the HTML / JavaScript code provided by professional page designers at TIF.

The presentation language also includes the possibility to express conditions on user profiles, the access device, the state of the application, or the characteristics of the information itself to be presented. This way, any aspect of presentation can dynamically adapt itself to the execution context. These conditions can determine the choice of one or other presentation model for an instance, or at a more detailed level, establish the aspect of small parts of the presentation, the inclusion or not of certain information fragments, the generation of hyperlinks, or the selection of one or other page component (lists, tables, trees, etc.).

Currently three presentation models have been defined for the implemented application: extended view, to show instances with maximum detail in a page; summary view, to show lists of instances, for example the ones that result from a search; and minimum view, to be used for example as the text of the link to an instance. Fig. 4. shows an example of an extended view for an instance of the Fundamental Analysis concept.

The explicit ontology allows describing more meaningful and precise user profiles, which can express preferences on specific topics, content classes, or even abstract content characterizations. The user models defined to this date include a) professional profiles, and b) subscription profiles of content consumers. The subscription profile carries access permissions to different parts of the ontology, as a consequence of which the user will have hidden access to different information areas. The professional profile defines a scale of interests for different subjects and types of materials, which determines the order (priority) and amount of information that is shown to the user, depending on the typology and relevant subject areas for his/her profile.

Fig. 4. Extended view of a Fundamental Analysis instance

3.5 Content management and provision tools

Content managers themselves are actually 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. These search facilities are provided as a complement of the content provision tools for content managers.

The tools for inputting contents, currently in use at TIF, have been adapted to allow defining richer semantics in terms of the ontology. In addition to filling forms with fields for instance properties, managers can create rich interrelations among contents or to external entities. The user interface for content managers is based on web forms that are generated automatically according to the content class. These data input forms are created by the system exactly the same way as search forms (see section 3.3): a) there is a default mechanism that takes into account property types to generate appropriate input controls, and b) one can instead define an input form template for each class.

The main difference is that content input and content search requirements usually need differently designed forms, therefore designers should create two different templates for each class accordingly. For example, not all class attributes need to be used for search, but it is likely that they all need to be provided values when a new instance is created by a content manager. Likewise, the set of fields that should be exposed when instances are shown to the end information consumer need not be the same as (typically they are a superset of) the ones that appear in a search form, and may be a subset of all fields required by an instance creation form.

4 Experiences and results

Our first observation is that at the time of this writing no proper ontology was available for the description of economic and financial information. Most standards for information exchange (most based in XML) in the field are specifically oriented to business processes and transactions, and only have small descriptions for economic data, rigid forms, or content packaging information (title, author, source, time stamps, etc.). Only a few provide extensive enough vocabularies for dealing with semi-structured, semantically rich information contained in documents like the ones TIF produces. NewsML [14] provides the IPTC Subject Reference System, a thematic taxonomy that includes a section for economy, but with much too broad terms for highly specialized financial analysts like TIF professionals. RIXML [15] provides or adopts several controlled vocabularies for aspects like subject, industrial sector, intended audience, and geographic location.

These standards are difficult to adopt from the beginning because of the particularities and specialization of the provider's business, and the inevitable regional bias. For instance, the GICS industrial sector subdivision standard adopted by RIXML considers the shoe industry as a single sector, while in Spain, the shoe sector being a highly developed industry, a finer subcategorization is desirable, e.g. distinguishing sports shoes, sandals, boots, men / women shoes, etc. Rather than integrating the standards, we have developed or own taxonomies, and we are currently defining export/import mappings to standards.

Besides contributing an ontology for a domain where no proper ontology had been defined before, our work has motivated a major revision and improvement of the existing categorization taxonomy used at TIF, which had been incrementally built on-demand over the years, without a clear a-priori evolution plan. A cleaner, more consistent and better organized classification scheme, and a better and clearer understanding by the industrial partner of its own domain, has resulted from this project.

A second observation is that whereas Semantic Web technologies have reached a significant maturity level, we still miss certain tools or features that we felt should be basic. For instance, we are not aware of any freely available tool to dynamically link ontology instance properties to database records so that data are retrieved at runtime, or even instances are created from data on demand. Instead, we are using a mechanism to statically dump the whole database to create a huge RDF graph with all possible instances, which is not optimal, and launch this mechanism every day to update the graph with new data. We have found other minor, though no less important, limitations in commonly used Semantic Web tools as well, like the lack of an operator for string comparison in RDQL/Jena.

Another small detail that required more attention and effort than expected is that of text normalization for search purposes, which requires conversion of texts to uppercase form, removal of accents, maintaining a record of lexical variants for nouns (e.g. "John Doe", "Doe, John", "J. Doe", "Doe, J.", "Doe", "J.D."), etc. We have not included support for spelling errors and typos.

A public demo of our system and further information about the project is currently available at http://nets.ii.uam.es/aniceto/. The total set of data and documents produced and stored by TIF since the old system was put to work in 1998 amounts to

159,429 records stored in different DB tables, taking 5.1 Gb disk space (including a number of documents in PDF format). This volume of information has given rise to 180,831 instances and 2,705,827 statements in the RDF knowledge base, taking 1.3 Gb in the MySQL Jena format. The current version of the ontology includes 196 classes and 99 properties.

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 [10]. The work presented here is a contribution in this direction, and provides a testing ground for our research.

We have developed a platform for economic and financial information management using state-of the art Semantic Web technologies and standards. The platform includes an ontology-driven knowledge base, where information products are enriched with semantic descriptions. The platform provides means for content provision, access, and administration of this knowledge repository. The information access facilities include semantic-based search, exploration and visualization facilities. The advantages of the search, visualization, and management modules do not lie only in their application to the particular case at hand. Besides improving the end-user experience, they provide important advantages for developers, as flexible, general-purpose modules, portable to other ontologies, easy to configure, supporting a variety of options and power vs. simplicity levels. We actually intend to prepare these modules to make them publicly available.

Many aspects of our work so far can be improved. For instance, while our ontology defines various relations between content classes, such relations are poorly described by current data at TIF. The new ontology-based content management tools allow and encourage interrelating contents, once our platform is deployed, but the information will still be missing in previously existing contents. Due to the size of the legacy materials (over 5 Gb), the most feasible way to enrich old contents would be through (supervised) text analysis and (semi)automatic metadata extraction. Contents managers would also benefit from such a service, which can be used to assist manual annotation, relieving managers from part of this effort.

Another major direction for our future work is to extend the adoption of Semantic Web technologies to package, integrate and commercialize financial services currently offered by TIF. We are starting an extensive analysis of current financial services in the company in order to identify basic (atomic) services, implement them as web services, and study potential compositions into more complex and added-valued web services by using ongoing business process definition languages. This work intends to include semantic description of the service functionalities in order to enable dynamic discovery of services and contingency plans in case of error. The semantic descriptions of the services will make use of the TIF economic and financial ontology defined in the work presented in this paper.

Acknowledgements

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

References

1. J. L. Alonso, C. Carranza, P. Castells, B. Foncillas, R. Lara, M. Rico. Semantic Web Technologies for Economic and Financial Information Management. 2nd International Semantic Web Conference (ISWC'03), Poster Session. Sanibel Island (Florida), 2003.
2. Berners-Lee, T., Handler, J., Lassila, O.: The Semantic Web, Scientific American, May 2001.
3. 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.
4. Coates, A. B.: The Role of XML in Finance. XML Conference & Exposition 2001. Orlando, Florida, December 2001.
5. Ding, Y., Fensel, D.: Ontology Library Systems. The key to successful Ontology Re-Use. In: Proceedings of the First Semantic Web Working Symposium. California, USA: Stanford University 2001; S. 93-112.
6. Gómez-Pérez, A., Angele, J., Bechhofer, S., Corcho, O., Domingue, J., Légér, A., Missikoff, M., Motta, E., Musen, M., Noy, N. F., Sure, Y., Taglino, F., McGuiness, D., Ramos, J. A., Stumme, G., Bouillon, Y., Fernández-López, M., Stutt, A., Handschuh, S., López, A., Maier-Collin, M., Christophides, V., Plexousakis, D., Magkanaraki, A., Ahn, T. T., Karvounarakis, G.: OntoWeb deliverable 1.3: A survey on ontology tools, available at http://www.ontoweb.org, 2002.
7. S. L. Green, S. J. Delvin, P. E. Cannata and L. M. Gómez. No Ifs, ANDs or Ors: A study of database querying. International Journal of Man-Machine Studies, 32 (3), pp. 303-326, 1990.
8. Gruber, T. R.: A translation approach to portable ontology specifications, Knowledge acquisition, 5(2), 1993.
9. K. Fast, F. Leise and M. Steckel. What Is A Controlled Vocabulary? Boxes and Arrows, December 2002. Available at http://www.boxesandarrows.com/archives/what_is_a_controlled_vocabulary.php.
10. S. Haustein and J. Pleumann. Is Participation in the Semantic Web too Difficult? International Semantic Web Conference (ISWC'2002). Sardinia, Italy, 2002.
11. M. Hearst, A. Elliott, J. English, R. Sinha, K. Swearingen, K-P. Yee, Finding the Flow in Web Site Search, Communications of the ACM, 45 (9), September 2002.
12. M. A. Hearst. User Interfaces and Visualization. In R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. Addison-Wesley, 1999, pp.257-323.
13. International Press Telecommunications Council (IPTC). Subject Reference System & NewsML Topicsets, http://www.iptc.org/metadata, 1999.
14. International Press Telecommunications Council (IPTC). NewsML, http://www.newsml.org, 2000.
15. Research Information Exchange Language, RIXML, http://www.rixml.org, 2001.
16. eXtensible Business Reporting Language, RBXL, http://www.xbrl.org, 1998.