SEMANTIC WEB AND ECONOMIC AND FINANCIAL INFORMATION MANAGEMENT

[1]N K PRASANNA [1]ANJANEYULU ANNA, [2]SHASI M

[1] Institute of Public Enterprise, Hyderabad, India
[2]Andhra University, Visakhapatnam, India

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 effective generation, gathering, integration, management and consumption of complex and big volumes of rich information. While the information integration problem is common to all business areas, it becomes apparent in the financial area, where information coming from multiple sources has to be homogenized in order to facilitate analysis and to provide a uniform view of the business e.g. via a balance scorecard. In addition, financial information is subject to official supervision e.g. by central banks. As a consequence, the required information has to be obtained, homogenized following the supervisor rules, and validated, guaranteeing its quality. In this paper, we present part of our experience in the use of semantic Web technologies for the management of economic and financial information and the problems encountered. In addition, we briefly introduce XBRL, a language that has gained a big momentum for business reporting, and compare the paths followed by the XBRL community and the semantic Web community.

Keywords: *Semantic Web, Economic Management, Financial Management*

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 [1]. While some business sectors make an extensive use of this kind of information e.g. credit institutions, insurance companies or investment firms, all other sectors also need to manage and analyze their own financial information and, furthermore, the quality of such management and analysis becomes a key business success factor. The volumes, complexity and value of economic and financial information make it a particularly interesting target for the application of novel information integration techniques. For this reason, an evaluation of what the semantic Web has to offer in this context can serve as a relevant feedback to the semantic Web community, and as a valuable information for businesses who might decide to promote or adopt semantic Web technologies.

Analyzing economic and financial information, providing value-added analysis of such information, and helping our customers to better manage and exploit their information assets are precisely our core activities. Tecnolog´ıa, Informaci´on y Finanzas (TIF) is part of Grupo Analistas, a corporation that generates high-quality economic and financial information (equity research notes, newsletters, analysis, sector reports, recommendations), and provides technology solutions for information consumers to access, manage, integrate, exploit and publish this information. 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. An important group of professionals and domain experts in the company is in charge of daily generating a wide range of economic and financial information, including economic, market, bank, and financial analysis, commercial fair reports, import/ export offers and news and manuals, among others. Banking consultancy is also part of our services, which includes the integration and analysis of heterogeneous information, and the provision of

management information services that support decision-making by offering a coherent and unified view of key financial data.
Due to the nature of our services, we have investigated semantic Web technologies as a potential means to improve the solutions we offer to our customers and to improve our internal information management and consumption processes.
In the following, we will discuss about Semantic Web (Section 2), Applying Ontologies to Economical and Financial Information Management (Section 3) and conduct a discussion around the eXtensible Business Reporting Language (XBRL)[1] (Section 4). The aim of the discussion is to compare the paths followed by the XBRL community and by the semantic Web community, trying to extract conclusions on what industry is missing from the semantic Web approach.

2. SEMANTIC WEB

As the internet grew in popularity, more and more information became available and also a major overload – anytime you searched on the internet you ended up with more than 1,00,000+ links that provide information on a subject whether those were relevant or not. The Semantic web technology is garnered to address this problem by trying to put a structure to the content on the web.

The semantic web requires a language in which information can be represented. This language should support

a. knowledge representation and reasoning,
b. the description of document content,
c. the exchange of the documents and the incorporated knowledge and
d. standardization.

The first two aspects demand adequate expressiveness. The last two aspects emphasize that the semantic web, like the web, should be a medium for the exchange of a wide variety of objects and thus allow for ease-of-use and for agreed upon protocols.
Semantic Web provides a common framework that allows data on the web to be defined, shared, reused and linked across application, enterprise, and community boundaries in a way that it can be used by machines – not just for display purposes, but for using it in various applications. It has brought structure to the meaningful content of web pages, creating a environment where it might be possible to carryout sophisticated tasks for users from the links between pages and resources. Semantic web technologies, such as RDF, OWL, Notation 3, Ontologies have allowed expression of semantic rich content in the web pages, using which the software application can connect various resource descriptions that appear in the pages, together through their Uniform Resource Identifier (URIs).

2.1 XML

XML (Extensible mark-up language) was designed as a language for mark-up or annotation of documents. An XML object is a labeled tree and consists of objects with attributes and values that can themselves be XML objects. Beside annotation for formatting, XML allows the definition of any kind of annotation, thus opening the way to annotation with ontologies and to use as data model for arbitrary information.
XML Schema allows the definition of grammer for valid XML documents, and the reference to "name spaces", sets of labels that can be accessed via the internet, XML can also be used as a scheme for structured databases. The value of an attribute can be text but it can also be an element of a limited set or a number. XML is only an abstract data format.
Tools have been developed for search and retrieval in XML trees. Tools can create formatted output from formatting annotations but in general any type of operation is possible. When tools are integrated in the web and can be called from outside they are called "services". This creates a very flexible representation formation that can be used to represent information that is partially structured.
XML helps organizing documents by providing a format syntax for annotation. Erdmann[2] provides a detailed analysis of the capabilities of XML, the shortcomings of XML concerning semantics and possible solutions. For Web Mining the standardization created by XML simplifies the development of generic systems that learn from data on the web.

2.2 RDF

The *Resource Description Framework (RDF)* is, according to the W3C recommendation [3], "a foundation for processing metadata; it provides interoperability between applications that exchange machine-understandable information on the Web". RDF is based on the idea of identifying things using web identifiers (URI) and describing resources in terms of sample properties and property values.

RDF documents consist of three types of entities: resources, properties and statements. Resources may be Web pages, parts or collections of Web pages, or any (real world) objects which are not directly part of the World Wide Web. In RDF, resources are always addressed by URIs, Universal Resource Identifiers, a generalisation of URLs that includes services besides locations. Properties are specific attributes, characteristics or relations describing resources. A resource together with a property having a value for that resource form an RDF statement. A value is either a literal, a resource, or another statement. Statements can thus be considered as object-attribute-value triples. Summarising, RDF and RDF schema provide base support for the specification of semantics and use the widespread XML as syntax. More information on RDF(s) can be found on the W3C website (www.w3.org) and many books.

2.3 OWL

Like RDF and RDF schema, OWL (Web Ontology Language) is a W3C recommendation, intended to support more elaborate semantics. OWL includes elements and attributes, which have a standard meaning and are used to define terms and their relationships. It was designed to provide a common way to process the content of web information as well as to be read by the computers. It comes with a larger vocabulary for describing properties and classes: among others, relations between classes, cardinality, equality, richer typing of properties and characteristics of properties and enumerated classes.

The expressiveness of OWL comes at a high cost. First, OWL contains constructs that make it undecidable. Second, reasoning is not efficient. Third, the expressiveness is achieved by increased complexity, so that ease-of-use and intuitiveness are no more given. OWL has three sub languages namely, OWL Lite, OWL DL and OWL FULL. OWL Lite supports those users primarily needing a classification hierarchy and simple constraints. OWL DL supports those users who want the maximum expressiveness while retaining computational completeness and decidability. OWL DL includes all OWL language constructs, but they can be used only under certain restrictions. OWL FULL is meant for users who want maximum expressiveness and the syntactic freedom of RDF with no computational guarantees. OWL Full is upward compatible with RDF. A legal OWL DL document is also a legal RDF document but not vice versa.

The development of OWL and its application is still in an early stage. If it leads of the availability of large knowledge bases via the internet, this will increase the relevance of knowledge-intensive Data Mining methods, that combine data with prior (OWL) knowledge.

For example http://webanalytics.com\prasweblog\R_TimePeriod.html shows the weblog report in the terms of Daily, Monthly, Weekly we can write it as

http://webanalytics.com\prasweblog\R_TimePeriod.html#ByDay

http://webanalytics.com\prasweblog\R_TimePeriod.html#ByWeek

http://webanalytics.com\prasweblog\R_TimePeriod.html#ByMonth

2.4 Notation 3 (N3)

N3 is basically equivalent to RDF in its XML syntax, but is more compact, readable and allows greater expressiveness. For example, the RDF description as shown above has the URI http://webanalytics.com\prasweblog\R_TimePeriod.html last part changing each time. N3 gives us an excellent way to abbreviate this – by giving parts of URIs, aliases, and using those aliases instead. Thus we can give an alias, say "pt" to the URI that occurs frequently, as in our case it is http://webanalytics.com\prasweblog\R_TimePeriod.html, and then use that to describe the terms like ByDay, ByWeek, ByMonth. For example, we could rewrite triples as

@prefix pt: <http://webanalytics.com\prasweblog\R_TimePeriod.html#>

Pt:ByDay pt:ByMonth pt:ByWeek

2.5 Closed World Machine (CWM)

CWM is a general-purpose rule engine for the semantic web. It is a forward chaining reasoner which can be used for querying, checking, transforming and filtering information. Its core language is RDF or N3, extended to include rules and it uses RDF/XML or RDF/N3 serializations as required. CWM takes in an RDF or N3 data file and a rule file describing rule son the RDF/N3 data file, as input and applies the rules to the RDF or N3 lexicon to generate all the RDF triples that can possibly be inferred on the data [4].

2.6 Ontologies

Beside the formal languages to be used for the SemanticWeb there is the ambition to develop ontologies for general use. There are in practice two types of ontologies. The first type uses a small number of relations between concepts, usually the subclass relation and sometimes the part-of relation. Popular and commonly used are ontologies of Web documents, such as DMoz or Yahoo!, where the documents are hierarchically organized based on the content. For each content topic, there is an ontology node, with more general topics placed higher in the hierarchy. For instance, one of the top level topics in DMoz is “Computers” that has as one of the subtopics “Data Formats. Under it, there is a subtopic “Markup Languages” that has “XML” as one of its subtopics. There are several hundred documents assigned to the node on “XML” or some of its sub nodes. [1]Each Web document is very briefly described and this description together with the hyperlink to the document is placed into one or more ontology nodes. For instance, one item in the “XML” node is a hyperlink to W3C page on XML, http://www.w3.org/XML/, with the associated brief description: “Extensible Markup Language (XML) - Main page for World Wide Web Consortium (W3C) XML activity and information”. We can say that here each concept (topic in this case) in the ontology is described by a set of Web documents and their corresponding short descriptions with hyperlinks. The only kind of relations that appear in such ontologies are implicit relations between more specific topic, that is a “subtopic of” a more general topic while the more general topic is a “super topic of” a more specific topic.
The other kind of ontologies are rich with relations but have a rather limited description of concepts consisting usually of a few words. A well known example of a general, manually constructed ontology is the semantic network WordNet [5] with 26 different relations (e.g., hypernym, synonym). For instance, concepts such as “bird” and “animal” are connected with the relation “is a kind of”, concepts “bird” and “wing” are connected with the relation “has part”.

[1] See http://dmoz.org/Computers/Data_Formats/markup_languages/XML/

3. APPLYING ONTOLOGIES TO ECONOMIC AND FINANCIAL INFORMATION MANAGEMENT

[1] represents a research project done in cooperation with Universidad Aut´onoma de Madrid (UAM)[2] to develop an ontology-based platform which adds explicit semantics to part of our information, and uses such semantics for visualization and search purposes.
As a result, with the development of ontologies, the annotation of existing content and the exploitation of the added annotations; details can be found in [1]. During the project it is found that some tools were not ready for industrial development, and efficiency was still an open problem.
One of the most important steps accomplished was the creation of a domain ontology from scratch, as no economic or financial ontology was available by that time. However, things have not changed much since then: ontologies that reflect the real consensus of entities producing and consuming information in the same domain are hardly available, which eliminates the possibility of expressing information in some agreed terms and, therefore, seriously reduces the benefits of using ontologies and semantic Web technologies.
The major conclusion of the project has been two-fold:

- Making a commonly used conceptualization of a domain explicit has some benefits, such as providing a reference for communication (both among persons and computers) and helping to improve data quality based on this conceptualization.

- The major promise of ontologies and the semantic Web has been making (semi)automatic interoperation possible. However, the biggest challenge has not been solved: facilitating the definition of real ontologies i.e. ontologies that really reflect the consensus of a critical mass of people so that they can be actually shared and reused.

4. XBRL: A LOST OPPORTUNITY OR A LESSON TO BE LEARNT?

In parallel to the semantic Web research, the eXtensible Business Reporting Language has

[2] http://www.uam.es/

gained momentum as a means of exchanging business reporting information. In the following, we briefly introduce XBRL, the acceptance it is gaining in the financial domain, and we relate it to semantic Web research.

4.1 XBRL in a nutshell

XBRL is a language that builds on top of XML and XML Schema to provide users with a standard format in which information can be exchanged, enabling the automatic extraction of information by software applications [6]. For that purpose, XBRL defines taxonomies, which provide the elements that will be used to describe information, and instances, which provide the real content of the elements defined.

Taxonomies make use of five different types of XLink linkbases, namely: definition linkbases, calculation linkbases, presentation linkbases, label linkbases and reference linkbases. The first three types contain different kinds of relations between elements, whereas the last two types contain documentation of elements. Definition links describe relations among concepts in a taxonomy, such as generalspecial relations, that provide information on what an element actually is e.g. the specialization of some other concept. Calculation linkbases provide information on how some elements are calculated in terms of some other elements, which can be exploited for data validation. Presentation linkbases contain relations such as parent-child that are exclusively used for presentation purposes e.g. a given element will be shown as the child of some other.

The last two types of links do not define relations among elements but document elements in a taxonomy. Label links provide labels in natural language with the purpose of facilitating the understanding of data by a human user. XBRL comes with multilinguality support and enables the user to associate labels in different languages to the same element. Reference links point to legal or other type of documentation that explains the meaning of a given taxonomy element.

Instances are then created according to the elements defined in a given set of taxonomies and linkbases, constituting what is called a Discoverable Taxonomy Set (DTS)[6], and therefore structured, documented and interpreted according to such DTS. This gives data an agreed meaning in a similar way ontologies do.

4.2 XBRL in the economic and financial domain

XBRL is intended to cover the needs of a particular domain, and it has evolved with the needs of such domain in mind. At the time of writing, XBRL is being promoted by public institutions such as the Committee of European Banking Supervisors (CEBS)[3], which includes high level representatives from the banking supervisors and central banks of the European Union. CEBS has promoted the creation of working groups that have the mission of defining XBRL taxonomies to be later adapted and used for the financial reporting that banks and other institutions have to submit periodically to the banking supervisors. One of these working groups is COREP[4], devoted to the creation of an initial taxonomy for common solvency ratio reporting in the context of the new Basel II capital agreement[5].

COREP has already led to the creation of initial taxonomies that will be reviewed when the directive concerning the new capital agreement of the EU is final, and later adapted (if necessary) by the national supervisors but keeping a common ground. Grupo Analistas, as a member of the COREP group, is contributing to the development of this initiative.

XBRL is not only being promoted by CEBS, but also by other public institutions such as CNMV[6], the supervisor of the Spanish stock market, and a number of taxonomies are already approved, such as the US GAAP taxonomies. In addition, it can be seen that important software vendors e.g. Fujitsu[7] have already paid important attention to XBRL and some have even published their financial data in XBRL e.g. SoftwareAG[8] or Microsoft[9].

4.3 XBRL and the Semantic Web

---

[3] http://www.c-ebs.org/

[4] http://www.corep.info/

[5] http://europa.eu.int/rapid/pressReleasesAction.do?reference=IP/04/899&language=en&guiLanguage=en

[6] http://www.cnmv.es/

[7] http://xml.fujitsu.com/xbrl/

[8] http://xbrl.softwareag.es/xbrlSageSite/index.html

[9] http://www.microsoft.com/msft/xbrlinfo.mspx

While XBRL is more domain specific than semantic Web technologies, they have a fairly similar goal: making the semantics of information explicit so that information exchange can be eased. Towards this goal, XBRL seems to have focused on a language heavily based on existing standards, adding as little as necessary to provide a usable language for business reporting, while the semantic Web community has proposed and evolved a number of (sometimes complementary and sometimes competing) standards to solve the semantic problem for any kind of domain, without committing to any particular application area.

Even though it might seem that the scope of XBRL is more limited than the scope of semantic Web research, XBRL is being more effective than semantic Web technologies on having a real impact. From our point of view, the major reason for this is that XBRL has made an effort to have an application domain and to identify the needs of technology users. If a domain is to be formalized and consensus is to be reached, that can never happen without having in place the entities and people that should reach that consensus. On that regard, the semantic Web has advanced in some particular topics but has forgotten its pre-requisite: finding shared conceptualizations i.e. ontologies.

Currently, really agreed, ready-to-use ontologies are hardly available, while XBRL is already succeeding on defining agreed taxonomies and on bringing together a critical mass of business experts, bank supervisors, technology providers and consultants to agree on taxonomies. Still, XBRL will face similar problems to the ones semantic Web researchers are already working on, such as taxonomy mapping and alignment, semi-automatic extraction of instances from different data sources, etc. Therefore, it might be worthwhile for the semantic Web community to explore the XBRL effort and to combine strengths.

5. CONCLUSIONS

The economic and financial domain is a particularly interesting domain for the application of novel techniques that could improve current information integration processes. TIF has experienced the application of such technologies to its business, realizing that the process of creating ontologies has hardly started in industry. On the other hand, XBRL is succeeding at attracting a critical mass of business actors, which is enabling the creation of agreed taxonomies that will be used in the near future, promoted by e.g. bank supervisors. A brief study of both lines of research can easily show that XBRL has been running closer to industry, while the semantic Web community has partly forgotten the pre-requisite to have a semantic Web: having ontologies. Therefore, we expect semantic Web research to be slightly redirected to work closer with industry and with other efforts such as XBRL that are working closer to the needs of technology users. Semantic technologies, if there are no ontologies providing the semantics to work with, will be of little use.

REFERENCES:

[1]. Pablo Castells, Borja Foncillas, and Rub´en Lara. Semantic web technologies for economic and financial information management. In *ESWS 2004*, Heraklion, Greece, May 2004.

[2]. Erdmann,M.: Ontologien zur konzeptuellen Modellierung der Semantik von XML. Isbn:3831126356, University of Karlshruhe (2001).

[3]. W3C: RDF/XML Syntax Specification (Revised). W3C recommendation, http://www.w3.org/TR/2004/REC-rdf-syntax-grammer-20040210/(2004)

[4]. C.Anantaram, “Semantic Web Technology-The next generation internet,” *Computer Society of India*, vol.30, issue 6, pp. 37-39, 2006.

[5]. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA (1998)

[6]. eXtensible Business Reporting Language. Technical report, XBRL International. http://www.xbrl.org.