RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS No. 147
Advanced Information Technologies for Management – AITM 2010 2010

Frederic Andres
National Institute of Informatics, Tokyo, Japan
e-mail: andres@nii.ac.jp

Rajkumar Kannan
Bishop Heber College (Autonomous), Trichy, India
e-mail: rajkumar@bhc.edu.in

XBRL TOPIC MAPS CHALLENGE FOR BUSINESS SEMANTIC MANAGEMENT

Abstract: This paper presents an approach to analyze XBRL documents using semantic web technologies. XBRL is an XML-based standard widely used in the exchange and representation of accounting and business information between different organizations. The proposed system organizes XBRL documents under the topic maps data model (ISO/IEC1325) in relationship with an ontology-based PSI (Published Subject Identifiers) server describing financial information domain. The system enables the use of TMQL queries over the knowledge system that facilitates the analysis of the financial information. Currently the system is being developed as a collective intelligence business decision-making support.

Keywords: topic maps, SaaS, knowledge modeling.

1. Introduction

XBRL (eXtensible Business Reporting Language) is an XML based standard developed by a not-for-profit international consortium of more than 450 organizations including regulators, government agencies and software vendors. XBRL is a language for the electronic communication of business and financial data that is set to revolutionize business reporting around the world. It provides major benefits in the preparation, analysis and communication of business information. It offers cost savings, greater efficiency and improved accuracy and reliability to all those involved in supplying or using financial data 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 a specific XBRL taxonomy. In this paper, we will concentrate on the way to manage financial data using the Topic Maps Data Model. Leaded by ISO, a number of semantic web technologies have appeared, like Topic Maps to describe resources using a graph model, TMCL to define financial

and business constraints, and TMQL, to define queries over Topic Maps graph-based data. One important aspect of these technologies is that it enables the composition of business information and exchange between organizations.

The remainder of the paper is structured as follows: section 2 briefly describes the accounting standardization using XBRL. Section 3 points out the key challenges of XBRL topic maps and business semantic management. Section 4 overviews the architecture of the proposed Topic Maps based system for the analysis of the accounting information in XBRL format. Finally, section 5 outlines the main conclusions of our research.

2. The XBRL standard overview

XBRL is a language for the electronic communication of business data, part of the standardization of accounting reports. It provides major benefits in the preparation, analysis and communication of financial information. This open standard facilitates business intelligence (BI) automation by enabling machine-to-machine communication and data processing for financial information with an eye towards cost reduction through the elimination of time consuming and error-prone human interaction. It offers cost savings, greater efficiency and improved accuracy and reliability to all those involved in supplying or using accounting information. It is one of a family of "XML" languages that is becoming a standard means of communicating information between businesses and on the Internet. XBRL has been developed by XBRL International, which is an international non-profit consortium [http://www.xbrl.org] of approximately 600 major companies, organizations and government agencies (as of March 2010).

3. The key challenges

This section encompasses the key challenges of XBRL topic maps and business semantic management. The most important challenge is the financial information retrieval aspect [Hwang et al. 2008] for decision support [Kotsiantis et al. 2009a, b; Schultewolter 2010]. XBRL Topic Maps enable search engine [Okamura et al. 2007] to provide satisfactory performance in both recall and precision. Topic Maps is an integration technology so it makes information subject-centric usable. We can associate Topic Maps with Linked data entity extraction [Cryans et al. 2010; Strötgen et al. 2010] based on Zemanta or on OpenCalais.

In addition, Topic Maps enables to add new semantics along XBRL information flow [Krieger et al. 2008] where information has a time life. XBRL Financial data is one of the most valuable information if it is fresh. The fresher a piece of XBRL data is, the more semantics it carries [Luo et al. 2009]. The distributed and fragmented nature of XBRL financial information impacts the generation of partial topic maps. Cleaning processing is needed as financial.

Another key challenge is the integration of XBRL for decision-making tools and problem-solving tools. It will help the finance function spend less effort on XBRL data wrangling, and more on higher order and faster decision support functions.

The Problem-solving Map generation for decision making support brings together XBRL information from different sources and structures scattered over the web. The architecture (see Figure 1) is based on a domain-based ontology representing the XBRL concepts/topics with specific properties as XBRL knowledge and relationships. Note that the problem is XBRL data, the mapping methods are from corporate finance, and the human problem solvers are experts in the fields of accounting and corporate finance. The problem solver refers to a person or group of persons who have knowledge about the XBRL financial problem along the WEB 2.0 XBRL model approach [Wenyi, Xu Tao 2008].

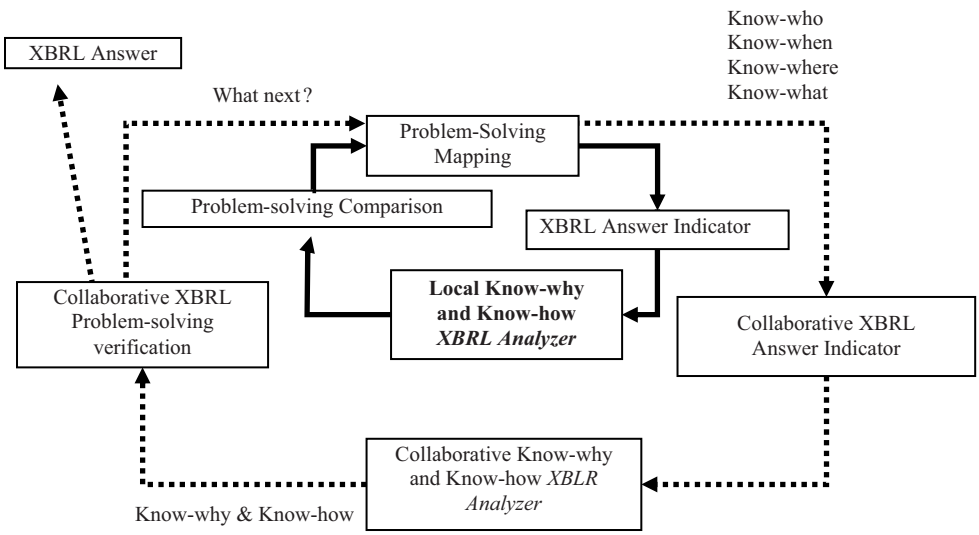

Figure 1. The XBR Problem-solving Mapping process

To create the problem-solving map generation, XBRL decision support is collected based on the collective intelligence procedure illustrated in Figure 1. The cycle drawn with the solid line corresponds to a local problem-solving mapping process involving a single user. The dotted cycle corresponds to a collaborative problem-solving mapping process involving several users.

Local Problem-solving Map process

Starting from the Problem-solving Map based on Information Extraction [Kawtrakul et al. 2007], it enables to determine the following XBRL knowledge inputs about know-why, and know-how. Answer indication annotates the XBRL knowledge input to be used by the local know-why and know-how XBRL Analyzer component. Then

the XBRL knowledge results, such as financial balances in order to improve the financial control process are verified by the Problem-Solving Comparison process.

Collaborative Problem-solving Map process

Starting from the *Problem-solving Map*, the collaborative answer indicator annotates the input with know-how and know-why XBRL information. The collaborative XBRL analyzer produces the XBRL knowledge results and the collaborative XBRL verification component evaluates the results.

4. The XBRLTM management system

The architecture of the XBRLTM system is depicted in Figure 2.

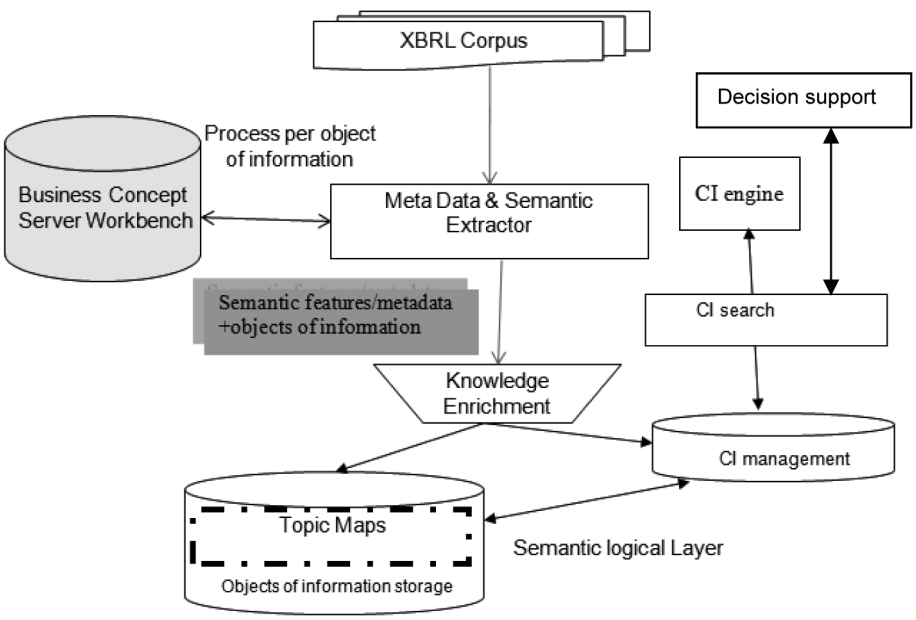

Figure 2. The XBRLTM system

The main components of XBRLTM system are:

- **Metadata & Semantic Extractor (M&S):** It takes the XBRL reports as input and converts them to Topic maps. Each XBRL concept is transformed into a topic.
- **CI encoding:** It combines the Topic Maps graphs obtained from using the ontology accounting principles (Business Concept workbench server). The Concept Workbench server defines over 100 concepts [Garcia, Gil 2010] from the Accounting Principles issued by the International Accounting Standards Board. The

results of the merger [Andres, Fukami 2007] are stored in the knowledge base of the system.

– Decision support and problem-solving XBRL Analyzer based on XBRL information extraction.

Collaborative XBRL analysis and annotation for XBRL knowledge extraction

The use of natural language processing extends the realm of semi-automatic XBRL knowledge acquisition and management. In addition, annotations are necessary to point out important information such as a collaborative solution to financial problems and guidelines for financial diagnosis.

Annotation types

The actual symbolic XBRL representations used in the corpus classify discourse annotations into two sets: causality tags for know-why XBRL extraction and procedural tags for know-how XBRL extraction.

Tagging means that the XBRL information is computer readable and can be more easily extracted, searched and analyzed by users of that information. We define causality tags and procedural tags.

Procedural tags support the collect of the XBRL knowledge for question answering system that responds to "how-to" questions involves unifying the question with a goal, and providing the user with the instructions associated with that goal. Procedural information will help to explain list of instructions to reach some financial goals. Procedural text includes documents as recommendations, directions for use/control/protection/prevention, and advice texts, etc.

The instructional structure of procedural financial reports is composed of the following items: (1) titles which express a goal to reach; (2) instruction compounds which express the financial actions to perform; (3) actions that are separated by two means: typographic and punctuation marks, for example bullet, new paragraph, etc., and linguistic marks which express temporality.

In procedural texts, titles express goals. To account for the hierarchy of tasks and subtasks, identified by titles (goals), we use an attribute that indicates the task level.

Collecting knowledge for a question answering system that replies to "how-to" questions involves unifying the question with a goal and providing the user with the instructions associated with that goal. Procedural texts provide instructions to reach a goal. They include documents such as formal recommendations, directions for use/control/protection/prevention, and advice.

Annotation process

We classify the annotation types into three independent layers: morphological, syntactic, and discourse. A semi-automatic analysis tool is used to annotate the XBRL linguistic information.

5. Conclusion and outlook

In this paper, we have presented the architecture of the XBRLTM system that facilitates the analysis of XBRL documents based on the Topic Map data model. The proposed SaaS system offers a great flexible approach for the accounting community and is adaptable to other domains and we are planning to apply it to various XBRL taxonomies.

Acknowledgements

This work has been supported by the National Institute of Informatics, Tokyo, Japan and the Bishop College (Autonomous), Trichy, India.

References

Andres F., Fukami N. (2007), Advanced semantic management of digital resources, [in:] *International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM'07)*, pp. 249-254.

Cryans J.M., Ratte S. Champagne R. (2010), Adaptation of apriori to MapReduce to build a warehouse of relations between named entities across the Web, [in:] *Proceedings of the 2010 Second International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA'10)*, IEEE Computer Society, Washington, DC, pp. 185-189.

Garcia R., Gil. R. (2010), Triplificating and linking XBRL financial data, [in:] *Proceedings of the 6th International Conference on Semantic* Systems *(I-SEMANTICS'10)*, A. Paschke, N. Henze, T. Pellegrini (Eds.), ACM, New York, Article 3, 8 pages. DOI=10.1145/1839707.1839711 http://doi.acm.org/10.1145/1839707.1839711.

Hwang J.S., Leem C.S., Moon H.J. (2008), A study on relationships among accounting transparency, accounting information transparency, and XBRL, [in:] *Third International Conference on Convergence and Hybrid Information Technology, ICCIT*, Vol. 1, pp. 502-509.

Kawtrakul A., Permpool T., Yingsaeree C., Andres F. (2007), A framework of NLP based information tracking and related knowledge organizing with topic maps, [in:] *NLDB2007*, Eds. Z. Kedad et al., *Lecture Notes in Computer Science*, Vol. 4592, Springer-Verlag, pp. 272-283, June 27-29, 2007, CNAM, Paris, France.

Kotsiantis S., Kanellopoulos D., Karioti V., Tampakas V. (2009a), On implementing an ontology-based portal for intelligent bankruptcy prediction, [in:] *2009 International Conference on Information and Financial Engineering 2009, ICIFE*, pp. 108-112.

Kotsiantis S.B., Kanellopoulos D., Karioti V., Tampakas V. (2009b), An ontology-based portal for credit risk analysis, [in:] *2nd IEEE International Conference on Computer Science and Information Technology 2009*, *ICCSIT*, pp.165-169.

Krieger H.U., Kiefer B., Declerck T. (2008), A hybrid reasoning architecture for Business Intelligence applications, [in:] *Eighth International Conference on Hybrid Intelligent Systems, HIS*, pp. 843--848.

Luo H., Gao J., Luo H. (2009), Web data extraction based on XBRL-GL taxonomy, [in:] *Asia-Pacific Conference on Information Processing, APCIP*, Vol. 1, pp. 358-361.

Okamura T., Fukami N., Robert C., Andres F. (2007), Digital resource semantic management of Islamic historical buildings case study on Isfahan Islamic architecture digital collection, *International Journal of Architectural Computing*, Vol. 5, No. 2, pp. 356-373.

Schultewolter C. (2010), Towards a framework for the generic specification of model-driven decision support systems: Classification criteria of model relationships, [in:] *43rd Hawaii International Conference on System Sciences HICSS*, pp. 1-10.

Strötgen J., Gertz M., Popov P. (2010), Extraction and exploration of spatio-temporal information in documents, [in:] *Proceedings of the 6th Workshop on Geographic Information Retrieval (GIR'10)*, ACM, New York.

Wenyi L., Xu Tao X. (2008), Research of WEB2.0-based XBRL network model, [in:] *2008 International Conference on Computer and Electrical Engineering, ICCEE*, pp. 209-213.

MAPY POJĘĆ XBRL ZARZĄDZAJĄCE SEMANTYKĄ BIZNESOWĄ

Streszczenie: W artykule przedstawiono podejście do analizy dokumentów XBRL przy użyciu technologii sieci semantycznych. XBRL jest oparty na języku XML standard powszechnie używanym w wymianie i reprezentacji rachunkowości oraz informacji biznesowych między różnymi organizacjami. Proponowany system organizuje dokumenty XBRL według modelu danych mapy pojęć (ISO/IEC1325) w powiązaniu z serwerem ontologii PSI opisującym informacje finansowe. System umożliwia korzystanie z kwerend TMQL do bazy wiedzy, co ułatwia analizę informacji finansowych. Obecnie system jest rozwijany jako kolektywna Business Intelligence wspierająca procesy podejmowanie decyzji.