THE ROLE OF THE SEMANTIC WEB IN STRUCTURING ORGANIZATIONAL KNOWLEDGE

Mangiuc Dragoş Marian
The Academy of Economic Studies in Bucharest Faculty of Accounting and Management Information Systems Bozieni 8, Bl. 831, Sc. A, Et.9, Ap.906, Sector 6, Bucureşti mangiuc@gmail.com Tel. 0723 – 22 78 76

The present paper is a component of an exploratory research project focused on discovering new ways to build, organize and consolidate organizational memory for an economic entity by means of the new "Semantic Web" technologies and also encloses some of the results of a previous doctoral research in the field of information technology assistance for the financial audit. The paper is an attempt to synthesize the ways organizational memory and organizational knowledge management may be optimized by means of the new "Semantic Web" philosophy. This paper offers a picture of where the typical organization is, and where it needs to be, in order to become a knowledge-aware organization and leverage the technologies of the "Semantic Web". As a consequence, a knowledge-centric process was defined, along with a "how-to" roadmap for crafting a company's way to the Semantic Web.

Keywords: organizational memory, organizational knowledge, semantic web, knowledge management

JEL code: M15

Introduction

In order to benefit from the "new wave" of semantic technologies, any modern organization must have a strategic view and also a reasonable understanding of the Semantic Web, XML, Web services, RDF, taxonomies, and ontologies. Each of these technologies has its distinct role in the build of organizational memory and the structure of organizational knowledge. It is the purpose of this paper to provide an opinion on how a company could be driven to take advantage of these technologies now so that it could start gather the rewards of the Semantic Web and prepare for the future. The paper focuses on three areas: diagnosing the problems of information management, providing an architectural vision for a modern organization, and providing some hints of how that vision may come true.

Research Methodology

The paper is a component of a wider research project called "Research in the Field of Modeling And Building Organizational Memory. *OMCAAF* – A New Methodological Framework for Financial and Accounting Cognitive *Acquis* Capitalization", and also continues a previous doctoral research in the field of computer-assisted financial audit tools and techniques, whose final results were publicly defended in order to be validated by the scientific and academic community. The main goal of the aforementioned research was the identification of some new areas of applicability for the modern knowledge-based information technologies in the field of financial audit.
In order to provide a set of valid and well-documented opinions about the realistic ways of augmenting the use of organizational memory by means of the modern information technologies, the author's proposals were preceded by an ample process of documentation and analysis of the field literature, allowing to get into terms with the main schools and opinion trends in the area, as well as the actual level of interconnection among the disciplines contributing to the present content of the "organizational memory", "organizational knowledge" and "Semantic Web" concepts. When possible, practitioners' expectations identification was attempted, both by means of questionnaires and direct interviews. In case some other author's opinion was enclosed, whether in exact quotation or synthetic form, a complete mention of the source identification information was made.
By defending the research results at the proceedings of such a scientific conference, attended by both scholars and practitioners bearing some interest in the research area, the author attempts to get further validation of his opinions, both confirmation and rejection of the aforementioned opinions' scientific and practical importance being welcome.

The usual organizational issue: too much information, too little knowledge

The most significant issue today for a typical organization is that information management is mostly chaotic. One important cause for the *status quo* is the huge amount of information coming in—from a wide variety of information sources (Berners-Lee, 1991). Making matters worse are the various formats of the data (paper, email, along with a wide variety of electronic media formats). Due to the amount of information coming in from various sources, its management gets more and more difficult. The "standard" organization is usually enclosing a lot of people getting overwhelmed with information (Fensel, 2007). Along with a missing cohesive information management vision, the typical organization has lots of information, but very little knowledge. The typical knowledge management process in an organization is depicted in the following figure (Figure 1).

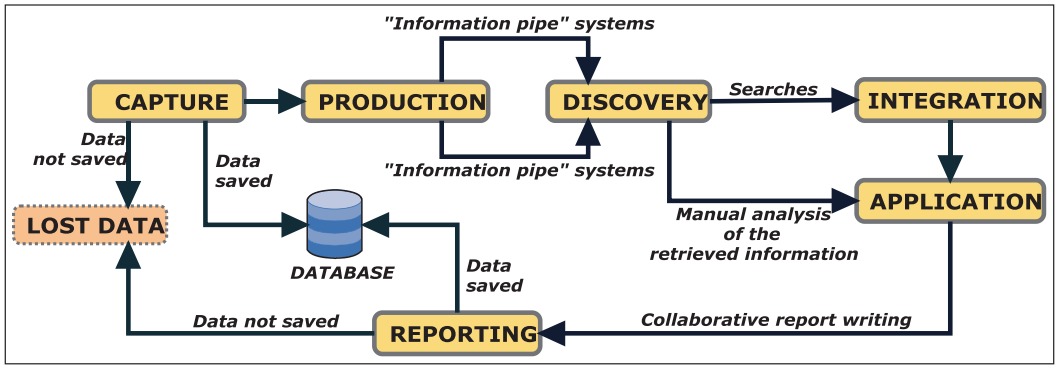

Figure 6 *– The typical knowledge management process*

The stages of this process usually are:

- The information capture – is the first stage in information management. Usually, an employee of the organization takes information from somewhere (newspaper, radio, Internet, database, phone call, customer contact, email) and brings it to the organization in some way. Many times, this is where the process stops (Ewalt, 2002). The employee may send it via email to someone, where it is lost in the "constellation" of emails that overwhelm the organization. If the data isn't lost in this way, a paper, a presentation, or a status report is usually written.

- The information production – is the second stage, (if information gets that far), where the data is put into a database, recorded to a digital file, or indexed into a search engine. Entering information is always the first step, but the problem is that each division, group, or project in the company enters the information into different systems. But a large-scale organization may enclose tens or even hundreds of different software systems dealing with information. Moreover, a financial database with the company's invoices, bills, and collections may add information to that total. Finally, the corporate human resources database must be taken into account. All the aforementioned software systems work like "information pipe" systems, able to perform a specific task, at the expense of "trapping" the data and reducing the organization's business agility and the capacity to adapt to new situations (Heflin *et al.*, 2002).

- The information integration – is the third stage, but it can also be missed, depending on the complexity of the organization's information architecture. Because of all the "information pipe" systems, there is usually no good way to gather all the information providers into a coherent picture. That is, any attempt to combine information in any way is a tiresome process, involving data conversions, incompatible software systems, and frustrated systems integrators (Patel-Schneider *et al.*, 2006). There is no repeatable process for integrating the systems, because each database and software system is designed differently and has different interfaces for communication purposes. As a result, there is usually little or no integration of the databases, because it is both prohibitively difficult and expensive. When there an integration solution arises, organizations usually pay a (very expensive) systems integrator to create a new and expensive "information pipe" system that integrates with the existing systems (Biezunski, 2003).

- The information search – is the fourth stage, depicting the process of "discovery" of the corporation's internal resources. This is usually random and time-consuming, involving many different systems. The user may have to log in to multiple databases and search engines, and manually compare and distinguish the information to get a big picture or coherent answer (Obrst and Liu, 2007). Even the results from search engines are usually based on keywords and Boolean logic, so they may or may not be relevant.

- The application of the search results - is the final stage of the process. After the tiresome search process, the result is usually a presentation or paper report. Many times, this process of creating the report involves several employees. The approval process is done by manual reviews and is slow. After the new product is created, the information may or may not be filed somewhere; it may be sent to a wrong or non-existent destination. If the report is filed, perhaps it is filed onto a Web server that may or may not be indexed by all (or even one) of the corporate search engines. If the new document is integrated into one of the corporate databases, there is no way to tell if the information has been superseded or is obsolete, which parts of the document are authoritative, and if the current version of the document has been approved by the organization. Lastly, there is information reuse – the ability to discover months or years later, and to refine, annotate, and incorporate past knowledge (Baader and Hollunder, 2001).

If any of these challenges are every day phenomena for an organization, then the organization may be in need of the Semantic Web. A well driven and well managed company will leverage the Semantic Web technologies to craft an information architecture vision, effective over every part of the organization life cycle.

The Knowledge-Centric Organization – A Big Step Ahead

A knowledge-centric organization will incorporate Semantic Web technologies into every part of the work life cycle, including production, presentation, analysis, dissemination, archiving, reuse, annotation, searches, and versioning. This section is a hint on how the semantic-oriented knowledge management process should be (see Figure 2):

Discovery and Production

The discovery and production phase is where an individual receives information and intends to translate it into organizational knowledge. The process should be an integral part of the corporate workflow process. This is an area where organizations should be aggressive an greedy in capturing information, because the effectiveness of reuse will be directly proportional to the quantity and quality of information captured. It is important that any new piece of information is marked up with XML, using a relevant corporate schema. Moreover, the individual should digitally sign the XML document using the XML Signature specification to provide strong assurance that the validity of the information has been verified. The annotation process may further arise, the employee may want to use RDF to annotate the new information with own notes or comments, adding to the XML document, but without breaking the digital signature seal of the original material. Finally, the author should digitally sign the annotation with XML signature. It is of main importance that before the information is integrated, its contents to be mapped to topics in the taxonomy and entities in the corporate ontology so that pieces of the information can be compared to other pieces of information in the corporate knowledge base. Once this is done, it is time to store the information in an application with a Web service interface, and any new Web service should be registered in the corporate registry, along with its taxonomic classifications.

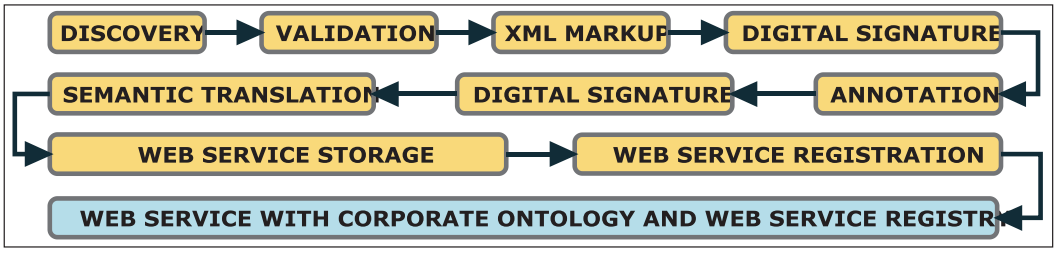

Figure 7 – *The semantic-oriented knowledge management process*

Search and Retrieval

As data is stored in an easily accessible format (Web services) and is associated with an ontology and a taxonomy, retrieval of information is much easier than the random process described in the previous section. Integration of all the Web services is not difficult because they all have a SOAP interface, and are registered in a corporate Web service registry; it is easy for an application to find what it is looking for (Berners-Lee *et al.*, 2001). Because of the hard work performed during the discovery and production process, the search and retrieval process is simpler and provides important functionality:

- Discovery of knowledge via taxonomies – this is the newfound power and possibly the killer app of the Semantic Web – the mining associations. In the field of electronic commerce, associations offer additional buying opportunities to customers.

- Pattern-based searches – all data can be semantically linked by relationships in the ontology, so patterns that would only be seen in the past by old data mining techniques that did not directly utilize meaning, can now be dynamically found with semantic searches.

- Manual and agent-based searches – although all of the searches can be manual, software agents can be endowed with rules to continually search the knowledge base and provide users with up-to-the-second results and alerts.

- Rule-based orchestration queries – Web services can be combined to provide modular functionality, so rules may be used to merge various searches from different Web services.

- Automated inference support – the corporate ontology explicitly represents concepts and their relationships in a logical and machine-interpretable form, so automated inference over the ontology and its knowledge bases becomes possible. Given a specific query, an ontology-based inference engine can perform deduction and other forms of automated reasoning to generate the possible implications of the query, thus returning much more meaningful results. The inference engine may discover inconsistencies or even contradictions in the ontology or knowledge bases (Wreder and Deng, 2007).

Application of Results

The final information production stage of the knowledge-centric organization's knowledge process is the application of results. If a new report has been created the responsible person should use the production process, depicted earlier in the "Discovery and Production" section. Another application in the last stage of the knowledge

process may be simple data annotation, requiring that the author of the annotation should digitally sign the annotation. Before the new annotation items are added, version control should be added to the document, and finally the document should be stored in the corporate information base. If an organization has a content management and workflow process that includes version control, annotation, and trust assertions, it will be easier to find information and apply the conclusions that were made earlier.

How could the "semantic" vision come true?

Most companies need to change their business process in order to take advantage of Semantic Web technologies. Luckily, these changes can be gradually implemented, and the organization can easily evolve into a knowledge-centric organization. The most challenging aspect may not be the technology; it may be changing the mind-set of the employees. Changing behavior and the ways that all levels think about accessing, integrating, and leveraging knowledge is critical:

How to Prepare for Change

The organization needs to be prepared. The stakeholders impacted by the change must be identified and led through the change process. A clear purpose and some clear goals and milestones are needed:

- A clear purpose for changing your information management process needs to be developed.
- Clear goals must be set.
- Stakeholders must be identified and a change plan must be developed.
- A core team that will help communicate the vision must be picked.

How to Begin Learning

In the author's opinion, a major time investment should be made in understanding the ideas and technologies behind the knowledge-centric process and all the implied technical staff must get to a reasonable level of knowledge. The following actions should be taken:

- The management staff must understand the main concept of the Semantic Web and the benefits of its adoption for the organization
- The technical staff to should be able to master the details of the soon-to-be-adopted technologies.

How to Create an Organization's Strategy

Now that management and the technical staff are "on board", it is time to design an organizational knowledge management strategy:

- Detailed technical goals must be set. For example:
- Corporate documents must be marked-up in XML.
- Applications should be exposed as Web services.
- Web services orchestration tools should be built.
- A corporate Web services registry should be established.
- Ontologies should be built.
- Tools that will help the production process should be used.
- Search tools should be integrated
- An enterprise portal should be used as a catalyst for knowledge engineering.
- A plan with a workflow change strategy must be developed.
- Appropriate staff must be set in place.
- A schedule must be set.

Conclusions

The present research is an attempt to identify the role of the Semantic Web in structuring knowledge at the organization level. The paper is an attempt to present an accurate view of the place where most of the typical large-scale organizations are today, and also the author's opinion on the place they need to be in order to become knowledge-centric organizations and leverage the technologies of the Semantic Web. As a consequence, a knowledge-centric process was defined and a "how-to" roadmap for crafting an organization's path to the Semantic Web was proposed.

Bibliography

1. Baader, F., Hollunder, B. – KRIS: Knowledge Representation and Inference System, Implemented Knowledge Representation and Reasoning Systems Journal, SIGART Bulletin 2, no. 3, June 2001
2. Berners-Lee, T. – What the Semantic Web Can Represent, September 1991, Available on-line at http://www.w3.org/DesignIssues/RDFnot.html
3. Berners-Lee, T., Hendler, J., Lassila, O. – The Semantic Web, The Scientific American, May 2001
4. Biezunski, M. – XML Topic Maps: Creating and Using Topic Maps for the Web, Addison-Wesley Publishing, 2003
5. Ewalt, D. – The Next Web, Information Week, October, 2002
6. Fensel, D. – Semantic Enabled Web Services, XML-Web Services ONE Conference, June 7, 2007
7. Heflin, J., Volz, R., Dale, J. – Requirements for a Web Ontology Language, W3C Working Draft, July 8, 2002
8. Obrst, L., Liu, H. – Knowledge Representation, Ontological Engineering, and Topic Maps, Addison-Wesley Publishing, 2007
9. Patel-Schneider, P. F., McGuinness, D. L., Brachman, R., Resnick, L. A., Borgida, A - The CLASSIC Knowledge Representation System: Guiding Principles and Implementation Rationale, Implemented Knowledge Representation and Reasoning Systems, SIGART Bulletin 5, no. 3, June 2006
10. Wreder, K., Deng, Y. – Architecture-Centered Enterprise System Development and Integration Based on Distributed Object Technology Standard, Institute of Electrical and Electronics Engineers, Inc., 2007