A Pragmatic Application of the Semantic Web Using **SemTalk**

Christian Fillies Semtation GmbH Falkensee Germany +49 173-8996801 cfillies@semtalk.com Gay Wood-Albrecht **Bonapart Solutions** Naperville, IL USA +1 630 881-7065

Frauke Weichardt Beratung im Netz Potsdam Germany +49 172-3171281

wood-albrecht@mindspring.com

fweichhardt@fweichhardt.de

Abstract

The Semantic Web is a new layer of the Internet that enables semantic representation of the contents of existing web pages. Using common ontologies, human users sketch out the most important facts in models that act as intelligent whiteboards. Once models are broadcasted to the Internet, new and intelligent search engines, "ambient" intelligent devices and agents would be able to exploit this knowledge network. [1]

The main idea of SemTalk is to empower end users to contribute to the Semantic Web by offering an easy to use MS Visio-based graphical editor to create RDF-like schema and workflows. Since the modeled data is found by Microsoft's Office XP SmartTags, users can benefit from these Semantic Webs as part of their daily work with other Microsoft Office products such as Word, Excel or Outlook.

SemTalk's graphically configurable meta model also extends the functionality of the Visio modeling tool because it makes it easy to configure Visio to different modeling worlds such as Business Engineering and CASE methodologies but also to these features can be applied to any other Visio drawings.

This paper presents two applied uses of this technology:

Ontology Project: Department-wide information modeling at the Main emphasis was on linguistic Credit Suisse Bank. standardization of terms. Based on a common central glossary, local knowledge management teams were able to develop specialized models for their decentralized departments. As part of the knowledge management process local glossaries were continually carried over into a common shared model.

Business Process Management Project: Distributed process modeling of the Bausparkasse Deutscher Ring, a German financial institution. Several groups of students from the Technical University FH Brandenburg explored how to develop and apply an industry-specific Semantic Web to Business Process Modeling.1

General Terms:

Documentation, Human Factors and Standardization

¹ Copyright is held by the authors/owners. WWW2002, May 7-11, 2002, Honolulu, Hawaii, USA ACM 1-5811-449-5/02/0005.

Key Words: Semantic Web, Business Process Modeling, Glossary and Ontologies

1. Introduction

Millions of people and thousands of applications are adding information to the internet / intranet on a daily basis. Rather than quickly accessing relevant information or automatically executing remote applications, time and productivity are lost in the search for information or in hardwiring transaction Technologies are needed that semantically connections. understand information requests to deliver desired information or that provide the services necessary to execute remote applications.

As a meta layer of the HTML Web the Semantic Web stores additional meta information about text. Similar to whiteboard files or frameworks, most relevant facts are sketched out in a model.

The Semantic Web is still in its initial stages. Enormous possibilities for further development can be seen from the increasing number of pages available about semantic webs. Even though concrete applications are still very rare, the definition of XML standards such as RDF, RDFS and DAML+OIL by the W3C suggest a growing interest. Therefore, it is likely that an ever-increasing number of Semantic Web applications will be seen in the near future.

Based on our early experiences, we predict that this new technology will spread first within the intranets of larger. distributed enterprises where there is a continuous demand to fine-tune Knowledge Management structures. Both the creation and fine-tuning of these knowledge structures are easily accomplished using Semantic Web technologies. The first step is to create a central vocabulary within an ontological context and to standardize processing concepts.

The main idea of SemTalk is to empower end users to contribute to the Semantic Web by offering an MS Office based graphical editor [2]. Based on an easy to use Microsoft Visio-based modeling tool, RDF Schemas are created. Following most of the other initial product offerings in this area, SemTalk is primarily focused on Knowledge Management applications rather than on intelligent machines which require a very detailed level of modeling.

SemTalk, using a Microsoft Visio front-end, offers an easy to use editor for semantic web ontologies and processes. Using a graphically configurable meta model, Visio is then adapted to different modeling worlds such as CASE Tools and organizational models. These models, with the help of Microsoft Office XP SmartTags, allow users to use semantic webs as byproducts of their daily work with other Microsoft Office products such as Word, Excel or Outlook.

This article describes two practical applications of Semantic Web technology.

The goal of the first project was to create a department-wide information model within Credit Suisse. Based on a common central ontology local knowledge management teams are able to develop specialized models for their decentralized departments.

The second project involved distributed process modeling of the *Bausparkasse Deutscher Ring*, a German financial institution. Several groups of students from the technical university FH Brandenburg explored how to develop and apply an industry-specific Semantic Web.

2 Technical Architecture

SemTalk is built on a RDFS-like XML data structure. Standard RDFS has been enriched by diagramming information and object oriented features like *methods* and *states*. Optimized structures for basic inferences such as inheritance and graph traversals are also included. There is an object engine providing a COM API to allow the engine to be used within MS Office products. Microsoft's Visio was selected as the graphical viewer because it is commonly used and because it is completely programmable. An object engine is used to define the semantic structures/ meta model for the existing Visio shapes. Shapes are graphically defined and rules are created to specify which shapes are allowed to be connected to each other.

SemTalk supplies the infrastructure necessary to define complete modeling methods inside Visio. Examples of commonly used modeling methods available in SemTalk are DAML, ERP and the BPM methods. SemTalk also contains interfaces to CASE tools such as Rational Rose and to Business Process Modeling Tools. There is a simple report generator for creating HTML tables as well as XSL for formatting. The new Ontoprise's Ontobroker will give users access to a powerful reasoning engine while modeling and while using ontologies within MS Office. [3]

3 Comprehensive Departmental Information Modeling at Credit Suisse

The project at Credit Suisse consisted of several workshops to create the basic repository for what was to become a growing visual glossary. This glossary is under consideration to be used as a basis for a knowledge management system. Workshop results were summarized in the form of conceptual models. These models were then published on the Credit Suisse Intranet.

3.1 Assumptions

Large enterprises have difficulty maintaining a common corporate language because of rapid technological change and the continual integration of smaller companies or departments into larger conglomerates. This is particularly true in the IT area where there is an abundance of different architecture descriptions, strategy papers and rapidly changing technology. The knowledge contained in documents is often strongly bound to the vocabulary of individuals, and is therefore difficult to consolidate. Homonyms, words having the same sounds but different meanings, cause additional problems. Even in the IT area synonyms are emerging that can have quite different meanings depending on the department.

3.2 Project Goals

Project goals were both linguistic standardization and to populate a central glossary that was to be used by people who were either designing or managing department-specific applications. The goal was not to establish central control or to mandate application selection; it was to create awareness of available terms and solutions used by local knowledge managers or members of the modeling team. In order to ensure that glossary usage became a permanent part of everyday practice, a general consciousness of usage scenarios for each term had to be produced. This can be most effectively accomplished by using SmartTags in Office XP or by using Babylon glossaries. (Babylon is an internet based translation and glossary tool with an installed base of 150 million copies.)

In this project an infrastructure and a base vocabulary was prepared based on information contained in 100 relevant documents. Glossaries and/or models needed to be represented in as flexible a way as possible and in a reusable format such as RDFS so that they can be imported as index structures into technical applications such as Document Management and Content Management systems. Similar applications are the automatic document classification system or Portals.

From the start of this project initial requirements demanded that the glossary be available in the Intranet in a form suitable for many different types of users. This meant that it was not acceptable to use complicated technical notations, e.g., UML diagrams.

It was hoped that this project would form the basis for the structure of future knowledge management systems. "Bootstrapping" such a system is always complex. If there is not enough content available, the system will not be used sufficiently and therefore would never begin to develop a life of its own. However a complete ontology of all objects existing in the enterprise is also not possible. The world is constantly changing and the language of the enterprise needs to reflect these changes, which implies that a company-wide glossary is never completed.

Success of the project depended on being able to publish a glossary with sufficient content and basic graphic definitions to encourage users to use and update the glossary as appropriate. This required technology that is easy to use and integrated with standard office applications. Similar to the creation and indexing of textual web pages, this is best done if the system appeals to the users need to participate in the process. Within

this scope of this project only the creation and modeling of a glossary were required.

3.3 Semantic Web as a Knowledge Management System

The glossary consists of terms with definition text and Synonym/homonym relationships. Properties and subClassOf relations are explicitly defined. In order to store information models flexibly, topic maps and RDFS are popular XML-based technologies.

SemTalk is used as graphic editor. With help from SemTalk and RDFS, the models can be saved as individual HTML web pages in the Intranet with all of their embedded hyperlinks. This type of the knowledge representation does not require central maintenance of the complete model, just a coordinated approval mechanism for core terms.

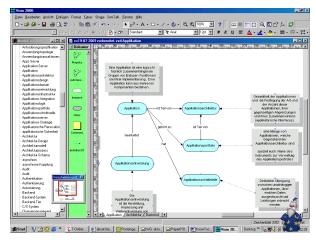


Figure 1: View of a SemTalk Model²

Consistency between different partial models is ensured during the modeling process by the SemTalk consistency Wizard. The Wizard points out which terms are already used in another model. Instead of modeling the same term again, a hyperlink to the reference term is inserted. The SemTalk Wizard uses index tables created by the SemTalk RDFS Crawler. This Crawler creates a directory of the available knowledge within selected areas of Intranet, Internet and within file systems.

These index tables are also used to interface with MS Office. SemTalk SmartTag is a technology that analyses text while the user is writing in order to mark the words that are already contained in the glossary as reference terms or Synonyms. Synonyms that are found can be replaced by reference terms if necessary. The definitions of the detected words are available using a single click that will take you to either the Visio model or to the available HTML representation. This results in substantial savings during complex manual revision of texts. Credit Suisse also uses glossaries created with SemTalk via Babylon.

The SemTalk Tool Suite points out documents and text passages to be revised. Specialized local models can be created as part of the document revision process. Models of individual documents or of specialty areas extend or add specialized components to the general glossary. As each term is used again it can be arranged in the context of existing terms. Queries of inference engines may reflect this subclass hierarchy. For example, if the general term "vehicle" is in the common glossary and "Porsche" is in the local document, you can search for "vehicle" and find your document about Porsche. If new terms for the general glossary emerge during document revision, they will be added after they are reviewed.

Knowledge management systems are often initially created via workshops, usually with expert interviews. Significant savings can be realized if the Concept composer from TextTech GmbH [6] is utilized to extract useful terminology.

- The Concept Composer is a text miner designed to search larger textual documents. Results are the most common terms and their collocation.
- Concept Presenter in the Intranet with graphic interface, can be integrated into the HTML Viewer of Semtalk.

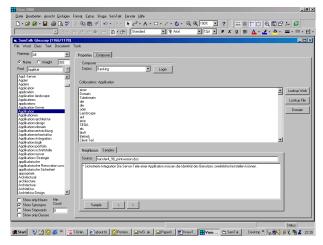


Figure 2: The Interface to Concept Composer

Different versions of definitions, associated synonyms, homonyms and text passages can be managed with the SemTalk Glossary. The SemTalk Glossary is the interface between SemTalk and the Concept Composer.

3.4 Project Bootstrapping

- Create a list of the most important terms
- Analyze text from 100 representative documents using the Concept Composer. Results are ranked by the importance of the technical terms. An infrastructure is created for looking up passages in the text and collocations that show the frequent word pairs. Concept Composer was used externally as ASP solution.
- Three, 3-5 days workshops, with up to five experts. During the workshops the SemTalk Glossary was used for the documentation and administration of definitions.

² This picture shows a part of the original project result which was done in German. You will find English samples at http://www.semtalk.com/semnet.htm

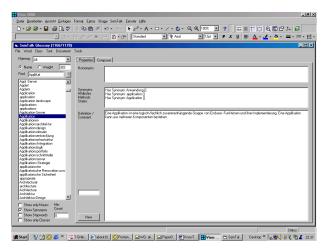


Figure 3: SemTalk Glossary

At the end of each Workshop day the scenarios discussed during the day are modeled graphically in SemTalk. The resulting graphic models are crucial in helping to simplify the following discussions. Relationships are easy to visualize and it is easy to navigate through large amounts of information. Homonyms are shown together in a picture to emphasize their different meanings.

At the end of the Workshop all central terms are defined and graphically modeled. The glossary with all of the graphic representations is then placed on the Intranet to be used by the enterprise.

3.5 The Knowledge Management Process

Creation of a glossary using SemTalk acts as a knowledge foundation that is designed to dynamically grow in ways that support better decision making and communication within the enterprise, especially as the environment changes. The glossary is published on the Intranet. Periodic audits of the contents ensure that the glossary remains up-to-date and useful. Modification requests are centrally collected and updates are made on a regular basis with the collaboration of the appropriate departments. Responsibility for the maintenance of the models was given to the individuals responsible for Intranet updates.

3.6 Project Results

Two hundred critical keywords were defined in seven workshops spread over a period of three months. Approximately 10 departmental representatives defined these keywords during the workshops that lasted between two hours and three days for each person. Two extra days for finishing the models were needed. Project costs were related to time lost from work. SemTalk Glossary was strongly felt to be a critical factor in being able to effectively build a glossary in such a short period of time.

The results were published in the Intranet and updated periodically. SemTalk enabled users to access keywords in several different contexts. The graphical view made it easier to understand the meaning of the keywords in relationship to each context because both the keyword and associated words are identified when doing searches.

SemTalk structured project work in a way that enhanced communication between coworkers from different departments. Additionally, purposeful revision of the documentation made it easy to quickly identify which documents needed to be updated, especially if context for a keyword changed.

3.7 Future Perspectives

The glossary created for Credit Suisse has been tested for more than six months. Acceptance is still high.

Future projects will significantly benefit from existing ontologies. Common IT related terms will hopefully be available as RDFS models on the Internet so that enterprise specific glossaries can further specialize those global structures.

4. Distributed Process Modeling at the Deutscher Ring Bausparkasse

The primary goal of this distributed process modeling project was to model *Order processing* at the Deutscher Ring Bausparkasse. This project took place over several weeks and was done by students from the University of Brandenburg.

The primary difference between this project and conventional process modeling projects was the use of an industry-specific Semantic Web. The Semantic web allows processes to be easily fine-tuned and terminological work to be executed more efficiently.

4.1 Conditions and Goals

Two separate groups, each with four students, modeled a business process in two different departments. After interviewing department members, information was modeled systematically in SemTalk. Models of existing processes were shown next to models of the "to be" processes that showed both the desires of each department as well as the feasibility of implementing the processes.

The primary customer targets were to make the processes clearer in the enterprise as well as defining the processes needed for the new workflow management system.

A significant project goal was to test the concept of distributed modeling in the context of the Semantic Web. The project team examined how communication can be improved within modeling teams and with the end-users.

Based on experiences in large modeling projects, effective distributed modeling requires more support from a tool than just providing a common repository. Even though such a repository can sometimes check the syntactic consistency of a model, more support is needed to create a common conceptual basis for functions, processes and information. This problem becomes more important if processes are spread between enterprises, e.g., such as the B2B area when different business partners must map their enterprise languages.

4.2. SemTalk Process Modeling Method

One of the most important philosophies behind the Internet, and hence Semantic Web, is that information is not copied, it is referenced. Creating links to external pages does not alter the contents of those pages. A flexible information system developed in this way does not have the consistency of a database but it has the advantage of being able to grow dynamically. SemTalk does not create individual models, it creates a network of linked models. While the emphasis of the Semantic Web is on pure knowledge representation, or in the case of Credit Suisse, the modeling of information classes, SemTalk process models can also be created and managed as a grid. Models can be linked with each other or they can be linked with external models such as models that represent industry-specific standards.

Semantic Web process modeling procedures consist primarily of three steps:

- 1. Selection of suitable reference libraries from the Internet
- 2. Customization of these libraries to fit project requirements
- 3. Creation of the process model using the reference model as a background

4.2.1. The Semantic Web Delivers Reference Models

Our methodology consists of using internet-based reference models that are easy to adapt to users needs. There is an increasing number of organizations that have developed such models:

- http://www.eccma.org is a large ontology which classifies services and products in order establish common understanding in E-business.
- http://www.dmtf.org develops an ontology for the Telecommunication Industry
- http://www.bpmi.org develops a process ontology for representing business processes
- http://www.papinet.org develops global transaction standards for the paper supply chain.
- http://www.hr-xml.org is dedicated to the development and promotion of standardized XML vocabularies for human resources (HR).

There are also different XML-based languages being used. Two popular repositories from the EAI area are BizTalk www.biztalk.org and RosettaNet.

General XML notation systems are found at www.cyc.com and at Wordnet www.xmlns.com

4.2.2 Process modeling

SemTalk supports different business process modeling methods, including the representation of enterprise processes named PROMET, a method developed by Österle at IMG (http://prometatweb.img.com/). In the current project, with its strong focus on internal processes, SemTalk uses the methodology of communication structural analysis (CSA) developed by Krallmann (http://www.sysedv.cs.tu-

berlin.de/Homepage/SYSEDV.nsf/). The students in the Deutscher Ring Bausparkasse project were already familiar with this method because of their experience with the CSA-based modeling tool Bonapart.

In CSA a process consists of interfaces between activities connected by information flows made up of information and media. Class models act as building blocks for these process models. Class models help to form structured and linguistic consistent process components. This improves re-use and allows object oriented reporting.

With SemTalk the class models in the Semantic Web are written in standard RDFS and they contain references to other class models. The class models can be created top-down using existing materials or bottom-up during workshops. Bottom-up modeling is generally more efficient because it helps to limit the modeling depth of the class models.

Thinking first about the objects and then over the processes themselves is an important step in the initial phases of the project. It is also critical to make sure that class libraries are consistent between several small related models. This will make it easier to integrate the models later.

4.2.3 An Example of Object Oriented Process Modeling

Address modification (Figures 4 & 5) is presented in the following example to demonstrate SemTalk's object oriented modeling method.

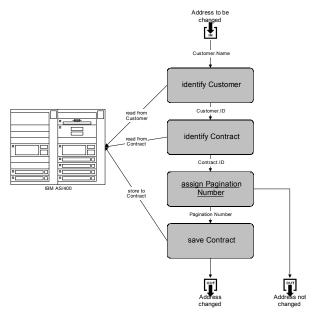


Figure 4: Example process Change Address

The key focus (beyond using Visio's shapes) is the naming consistency between tasks and associated business objects. The name of each task is a combination of the class name specified in the class model and a particular operation (verb) that is performed. Information flows may reference an attribute or a state. Object models are developed simultaneously as processes are being defined. Object model changes are immediately reflected in the process models. This technique allows the

creation of consistent and reusable process modules that can be used in even larger projects.

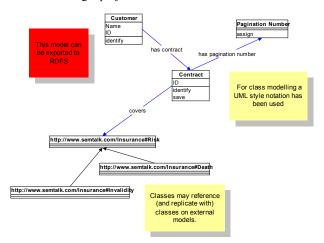


Figure 5: Class model for example process Change Address

The links to external models are further explained in the next section.

4.2.4 Tool Support for Distributed Models

SemTalk supports the user during the modeling process using Wizards that monitor the modeling process and offer Wizards are implemented as agents that suggestions. permanently check a given set of rules. Simple rules are tips about writing, e.g., upper/lower case, detecting synonyms and in the investigation of situations where the inheritance structure appears to be incorrect. The most important use for the Wizards is to find out whether the user is actually rebuilding models that already exist on the Semantic Web. Let assume the user is defining a class named "Vehicle". The Wizard will give him a hint that this concept has been already defined somewhere in another ontology such as ECCMA or WordNet. In this case the user should create a stub referencing the external definition of Vehicle as a hyperlink. Based on that hyperlink, the user's model can later be automatically updated once the definition of vehicle changes.

A class model can be linked to various RDFS data sources. Each class can be hyperlinked to a class in an external model. Single classes or complete models can be replicated from externally shared models. Although it is not part of the original intent of RDF, we use the same URN for encoding identity and location of a class.

The agents are supported by a Crawler, which looks independently or on request for available models and creates index files for the agents. The Crawler looks not only in the local file system but also in the Semantic Web for available sources of knowledge in the format RDFS.

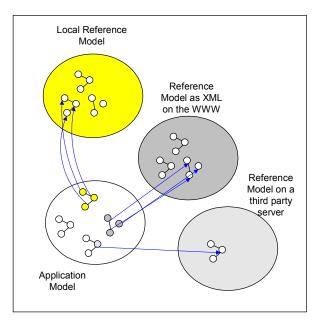


Figure 6: Hyperlinking SemTalk Models

4.3 Project Results

From customer's point of view this project was a success because it resulted in a concrete blueprint for workflow implementation.

The main difficulty for participants was the application of object oriented thinking to process modeling. This method significantly differs from the traditional way business processes are described.

5 Summary

SemTalk models give context to keywords. The Visio editor enables a wide range of users to use and understand models. The Visio editor helps to make modeling as simple and inexpensive as creating HTML web pages. This is a critical factor if the potential of the Semantic Web is to be achieved.

The addition of process modeling to the Semantic Web's class models broadens the reach of Semantic Web applications from **Ouality** Management to Process-Oriented Knowledge Management. It also helps to fill the gap between EAI and webbased services or E-Government. Using uniform, consistent, XML-based glossaries enterprises have new ways to share terminology between applications to ensure the meaningful integration of Content Management, Document Management and Data Warehouses solutions. Integrating SemTalk technology into daily work processes improves the acceptance, and thus the usefulness of the models. Finally, and most importantly, adding a process context unleashes the powerful and intelligent information retrieval possibilities offered by the Semantic Web.

6. References

- [1] Berners-Lee,T, Hendler,J. and Lassila, O. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities Scientifc American (May 2001)
- [2] Fillies, C.; Weichhardt, F.; SemTalk: A RDFS Editor for Visio 2000 Position Paper, ICCS 2001 9th International Conference on Conceptual Structures / Semantic Web Working Symposium (SWWS), Stanford Univ., July-Aug 2001
- [3] Staab, S., Studer, R., Schnurr, H.-P., Knowledge Processes and Ontologies. IEEE INTELLIGENT SYSTEMS, 1094-7167/01

- [4] Osterle, H., Vogler, P. Information Management Gesellschaft, PROMET, 1994; Business Engineering 1 1995, Praxis des Workflow-Managements, 1996
- [5] Krallmann, H., Feiten, L., Hoyer, R. & Kölzer, G.: Die Kommunikationsstrukturanalyse (KSA) - Zur Konzeption einer betrieblichen Kommunikationsarchitektur, Interaktive betriebswirtschaftliche Informations- und Kommunikationssysteme, Walter de Gruyter, Berlin, 1989
- [6] Heyer, G., Läuter, M., Quasthoff, U., Wittig, T. & Wolff, C., Learning Relations using Collocations. Proc. IJCAI Workshop on Ontology Learning, Seattle, WA, 19-24. August 2001