DATA INTEGRATION SOLUTION FOR PAPER INDUSTRY

A Semantic Storing Browsing and Annotation Mechanism for Online Fault Data

Sergiy Nikitin, Vagan Terziyan
Agora Center, University of Jyväskylä, Mattilanniemi 1, Jyväskylä, Finland

Jouni Pyötsiä
Metso Automation

Keywords: Semantic Web, Semantic Storage, Information Integration, Paper Industry.

Abstract: A lot of IT solutions exist for simplification and time saving of industrial experts' activities. However, due to large diversity of tools and case-by-case software development strategy, big industrial companies are looking for an efficient and viable information integration solution. The companies have realized the need for an integrated environment, where information is ready for extraction and sophisticated querying. We present here a semantic web-based solution for logging and annotating online fault data, which is designed, and implemented for a particular business case of a leading paper machinery maintenance and automation company.

1 INTRODUCTION

Rapid changes and discontinuities in the 21st century business environment will challenge companies with the growing demand for competitive advantages within their business solutions. To ensure high flexibility, sustainable growth and profitability, companies have to search for new innovative approaches to products and services development. New innovative business solutions call for strong integration of automation technology, information and communication technology (ICT), and business processes. At the same time, embedded intelligence in different machines and systems gives new possibilities for automated business process operation over the network throughout the machines and systems life cycles.

The new emerging remote service solutions imply that products transform into life cycle services and these services, in turn, transform into customers' service processes. Business messages coming from intelligent machines and systems drive these processes, utilizing embedded intelligence and ICT solutions.

In the future, a variety of collaborative resources, like intelligent machines, systems and experts, will create a huge amount of new information during the life cycles of machines and systems. Message flow management and compression to on-line knowledge are already a demanding issue for the logging of product-related activities. On the other hand, optimization requirements demand more effective knowledge utilization and the speeding up of network-based learning in the process of collaboration between different resources.

Industry challenges the IT-sector with the new requirements that are dictated by the need to offer essentially new services to customers in order to be competitive in the market. These requirements may become hard to meet using conventional tools and approaches. The growth in the information volumes we want to store and process by integrating data from different sources leads to an unprecedented level of complexity. Modern Enterprise Resource Planning (ERP) systems are trying to provide integrated solutions for large companies. However the installation and adjustment of such systems may take a half a year, involving hundreds of consultants and subcontractors.

The current trend towards more open service-based computing environments is a new approach to componentization and components distribution. Service-oriented architecture aims to achieve a new

Nikitin S., Terziyan V. and Pyötsiä J. (2007).
DATA INTEGRATION SOLUTION FOR PAPER INDUSTRY - A Semantic Storing Browsing and Annotation Mechanism for Online Fault Data.
In *Proceedings of the Fourth International Conference on Informatics in Control, Automation and Robotics*, pages 191-194
DOI: 10.5220/0001621801910194

level of reusability and business process flexibility; however, to ensure the interoperability between the components we need a common "glue" that would adjust the semantics of the data being exchanged. The Semantic Web technology (Berners-Lee, T., et al., 2001) introduces a set of standards and languages for representation of a domain model with the explicit semantics. The main instrument of domain model construction is ontology, which allows for domain data representation in a formalized and unified way.

In this paper we present a solution that utilizes Semantic Web technology to provide a tool for online maintenance data browsing, analysis and annotation. We utilize experience obtained in the SmartResource project (SmartResource, 2006) and apply the General Adaptation Framework (Kaykova et al., 2005) to align data with the domain ontology. The paper is organized as follows: In the next section we describe the paper machinery ICT infrastructure, and Section 3 presents the solution we have developed using Semantic Web tools and standards. We end with conclusions and future work.

2 IT INFRASTRUCTURE IN PAPER INDUSTRY

Metso Corporation is a global supplier of process industry machinery and systems as well as know-how and aftermarket services. The corporation's core businesses are fiber and paper technology, rock and minerals processing, and automation and control technology. Metso's strategy is based on an in-depth knowledge of its customers' core processes, close integration of automation and ICT, and a large installed base of machines and equipment. Metso's goal is to transform into a long-term partner for its customers. Based on the remote service infrastructure, it develops solutions and services to improve efficiency, usability and quality of customers' production processes throughout their entire life cycles.

2.1 Remote Service Infrastructure

Metso's remote service infrastructure consists of a service provider's Central Hub and several customers' Site Hubs, which are integrated over the network (see Figure 1).

The key issues in a Site Hub solution are: open standards, information security, reliability, connectivity and manageability.

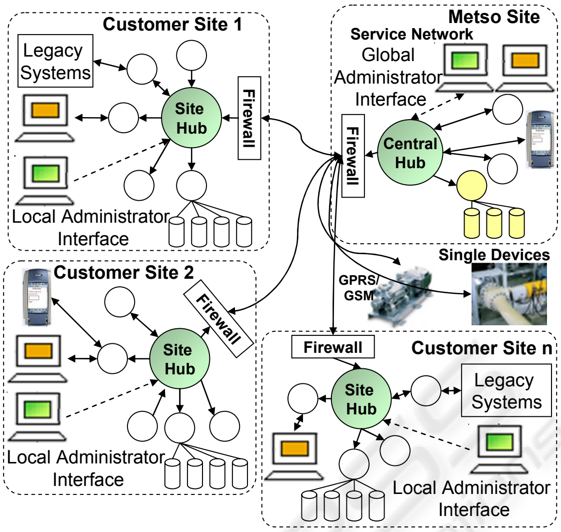

Figure 1: Site Hub network architecture.

Message Center is the main component of the Site Hub. It checks the validity of messages and routes them to the correct receivers. Messaging in Site Hub is based on Web Services technology.

Hub-based integrated infrastructure combined with secure connectivity allows easy incorporation of new business logic on both customer and Metso sites. A messaging mechanism between customers and Metso provides a very flexible medium for information exchange and new service provisioning.

3 LOGGING AND ANNOTATION OF MAINTENANCE DATA

The main purpose of the system we present here is to store alarm data, generated by paper machine monitoring systems. When an alarm happens, a SOAP/XML message (SOAP, 2003) is generated and sent to the Site Hub, which then forwards it to the Central Hub. We have established a message flow from the Central Hub to the computer at the university network, where messages are processed by our system.

3.1 Architecture of the System

The system can be divided into two main subcomponents – *Message Handler* and *Message Browser* (see Figure 2).

Message Handler receives and processes SOAP/XML messages from customers. It invokes the *Adapter* to transform the XML content into an

RDF-graph (RDF, 2004) object and store it in the Sesame RDF storage (Sesame).

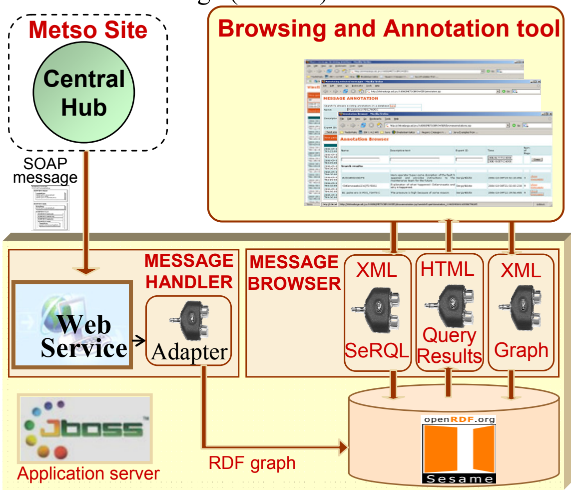

Figure 2: Architecture of the system.

The RDF storage contains an *Ontology* that plays the role of a schema for all data within the storage. Based on the analysis of SOAP/XML messages, we have defined main concepts (classes) with the corresponding properties (see Figure 3).

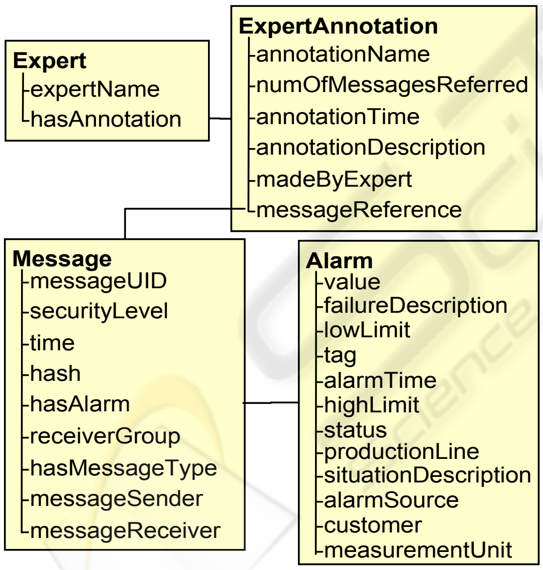

Figure 3: Ontology classes.

The *Message* class describes such message properties as message sender and receiver, message reception time, etc. The *Message* class also refers to an *Alarm* class, which contains information about the reason for message generation, such as measurements of sensors, status data and exact module of the production line where the alarm happened. The *ExpertAnnotation* class defines the structure for labelling groups of messages with an expert's decision and has references to instances of the *Message* class.

The *Message Browser* component provides a web-based interface for browsing and filtering messages stored in the RDF-storage, according to user-defined filtering criteria.

The purpose of message filtering is to distinguish the groups of messages leading to exceptional situations. The expert provides annotations for message groups which are stored to the RDF-storage and that can be used as samples for machine learning algorithms. The client-server interaction is implemented using AJAX technology (Garrett, 2005), which provides a more dynamic script-based interaction with the server (see Figure 4). When a user performs any action that requires invocation of server functionality, the script on a client side wraps the required parameters into XML format and sends it to the server. For example, in order to filter the messages, a user selects the needed parameters and specifies parameter values within the corresponding textboxes (see Figure 4).

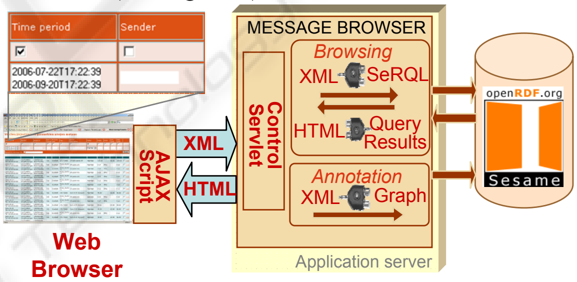

Figure 4: Client-server interaction.

On the server side, the *Control Servlet* handles the XML document. For filtering, it generates a SeRQL query (Broekstra, 2004) and executes it. On the client side, a dedicated callback script function processes the response and shows the result in a web browser.

3.2 Integration with the Agent Platform

We realize that the extension of the system will challenge the complexity of development and maintenance. That is why, following the autonomic computing paradigm (Kephart, 2003), we have tested agent-based scenario (see Figure 5) implemented on a JADE agent platform (Bellifemine; 2001). We have assigned an agent to manage RDF-storage activities (*Metso Storage Agent*) and provided a *Metso Expert Agent* to interact with a maintenance expert.

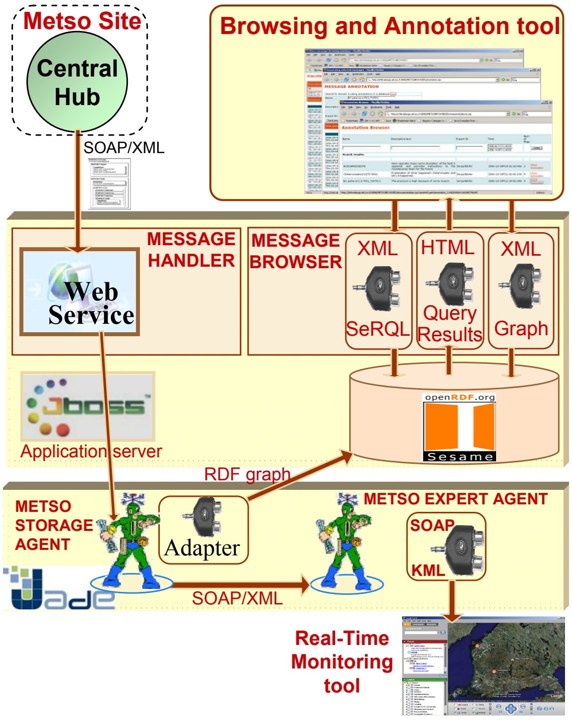

Figure 5: Agent-enabled system.

The messages coming from customers are handled by the *Metso Storage Agent*, which incorporates *Adapter* to perform transformation and storage. Then, the *Metso Storage Agent* sends the message to the *Metso Expert Agent*, which updates the situation on a Real-time Monitoring Tool and provides an expert with the message content and a link to the browsing and annotation tool.

4 CONCLUSIONS

Although we have succeeded with the implementation of the solution presented here, there are still many issues to cope with in order to meet key industrial requirements, such as scalability, maintainability and robustness. RDF-storages can handle billions of triples, but there are no mature semantic storage-oriented development patterns or guidelines. Nevertheless, the simplicity and efficiency of querying, as well as model extending, provide incontestable arguments in favour of semantic data storages. The ontological domain model brings more benefits to customers when there are more sources integrated. However, the complexity of such a system, if developed using conventional approaches, will be too burdensome to maintain and extend. In order to distribute the complexity, we introduce self-manageable entities in the agent-based communication scenario.

ACKNOWLEDGEMENTS

This research has been supported by the SmartResource project, funded by TEKES, and the industrial consortium of Metso Automation, TeliaSonera and ABB. The preparation of this paper was partially funded by the COMAS graduate school.

REFERENCES

Bellifemine, F., Poggi, A., and Rimassa, G. 2001. JADE: a FIPA2000 compliant agent development environment. In Proceedings of the Fifth international Conference on Autonomous Agents (Montreal, Quebec, Canada). AGENTS '01. ACM Press, New York, NY, 216-217. DOI= http://doi.acm.org/10.1145/375735.376120

Berners-Lee, T., Hendler, J., and Lassila, O. (2001) The Semantic Web, Scientific American, Vol. 284, No. 5, pp. 34-43.

Broekstra, J., Kampman A., and F. van Harmelen. Sesame: An Architecture for Storing and Querying RDF Data and Schema Information. In D. Fensel, J. Hendler, H. Lieberman, and W.Wahlster, editors, Semantics for the WWW. MIT Press, 2001.

Broekstra, J., Kampman, A. SeRQL: An RDF query and transformation language. In Proceedings of the International Semantic Web Conference, ISWC 2004, Hiroshima, Japan.

Garrett, J., Ajax: A New Approach to Web Applications (white paper).,http://www.adaptivepath.com/publications/essays/archives/000385.php, February 2005.

Kaykova O., Khriyenko O., Kovtun D., Naumenko A., Terziyan V., Zharko A., General Adaption Framework: Enabling Interoperability for Industrial Web Resources, In: International Journal on Semantic Web and Information Systems, Idea Group, ISSN: 1552-6283, Vol. 1, No. 3, July-September 2005, pp.31-63.

Kephart J.O., Chess D.M., 2003. The vision of autonomic computing, IEEE Computer, Vol. 36, No. 1, pp. 41-50

RDF – Resource Description Framework, A W3C Recommendation, Feb 2004, http://www.w3.org/RDF/

SmartResource – a TEKES funded project, http://www.cs.jyu.fi/ai/OntoGroup/SmartResource_details.htm

SOAP – Simple Object Access Protocol, A W3C Recommendation, 2003, http://www.w3.org/TR/soap/