SEMANTIC WEB SUPPORT FOR BUSINESS PROCESSES

Airi Salminen, Maiju Virtanen
University of Jyväskylä, PL 35 (Agora), Jyväskylä, Finland
Email: airi.salminen@jyu.fi, maiju.virtanen@jyu.fi

Keywords: Business processes, semantic web, RDF, RDF Schema

Abstract: Development of semantic web technologies has been initiated to improve the utilization of web resources particularly by software applications. Semantic web is intended to extend the current web by metadata adding meaning to web resources. In an interorganizational business process context, semantic web could be an extension of the current intranet, extranet, and internet resources better enabling computers and people in business processes to work in cooperation. In the paper we will explore the possibilities of the semantic web technologies to support business processes. Particularly we will evaluate the possibilities and problems related to the utilization of RDF (Resource Description Framework), a method supporting the formal presentation of metadata and metadata schemas. We will use the Finnish legislative process as a case to demonstrate the issues discussed.

1 INTRODUCTION

The wide adoption of the information and communication technology innovations and web technologies from 1990's has lead into a situation where a number of software applications and enormous data resources are available in business processes. Utilization of the resources in the processes, however, still requires vastly human work.

The World Wide Web Consortium (W3C) has initiated the development of semantic web to improve the utilization of web resources. The *semantic web* is intended to be "an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation" (Berners-Lee, Hendler, Lassila, 2001). The well-defined meaning is added to the web by means of metadata. The *metadata* is information about resources either accessible or identifiable on the web. The metadata is given in a formal, standardized format, readable and interpretable by software. The Extensible Markup Language (XML) and the Resource Description Framework (RDF) create the basis for the standard formats (Bray, Paoli, Sperberg-McQueen, Maler & Yergeau, 2004; Manola & Miller, 2004). Formal presentation of the metadata can be used to facilitate automated reasoning about the meaning and trustworthiness of resources. Ontologies are used to express semantic metadata. An *ontology* defines formally the concepts and their relationships in an application domain (Gruninger & Lee, 2002; Klein, 2002).

In the paper our goal is to explore the possibilities of the semantic web technologies to support business processes. Particularly we will evaluate the possibilities and problems related to the utilization of RDF. By the term *business process* we refer to work processes in all kinds of organizations, in companies, public sector organizations, as well as in other types of organizations. In the paper we will use the legislative process as a case to demonstrate the issues discussed.

The Finnish legislative process is an example of a complex interorganizational process participated by many organizations, among them the Government, the ministry on the domain of the law, the Ministry of Justice, the Chancellor of Justice, the Parliament of Finland, Special Committees of the Parliament, and the President of the Republic. The time for developing a new law may take from a couple of months to several years. A number of intranet systems, extranet systems, and various software applications are used during the process. Heterogeneity of systems in the participating organizations and lack of communication between software applications cause lot of extra work in the process. The major requirements for the adoption of new information and communication technology solutions include their capability to support

- data integration and interoperability of systems in the process, and

- building of intelligent services to the heterogenious group of users.

Semantic web technologies seem to provide, at least in principle, tools for the needs.

The rest of the paper is organized as follows. Section 2 first gives an overview of our ideas about semantic web in the context of a business process. Section 3 will then briefly describe RDF and RDF Schema. The opportunities offered by RDF/RDFS and problems related to the definition of RDF schemas will be discussed in Sections 4 and 5, respectively. Before the concluding section implications of the study for building semantic webs for business processes will be discussed.

2 TOWARDS SEMANTIC WEB TO SUPPORT BUSINESS PROCESSES

A semantic web supporting business processes involving a group of organizations would be an extension of their current intranet, extranet, and internet resources. A metadata standard should be developed to express the well-defined meaning of information used and produced in the processes. Adding well-defined meaning to the information in a standardized form is a challenging task.

In a business process context, the meanings are created and shared by the human actors in the process. There are meanings, for example, in the words and expressions of the texts in documents, in figures, in the organizational culture, in the gestures of people, in the ways the activities proceed, coded in software, and built in hardware.

To define formally the shared concepts and the actual metadata descriptions in a business process, a flexible description language is needed**.** It should facilitate description of the common concepts related to the process context of resources by means of *contextual metadata*, the description of the concepts related to the meaning of the resources by means of *semantic metadata*, as well as description of the physical or logical structure of the resources by means of *structural metadata***.** The importance of metadata about the process context of recorded data is pointed out also in the ISO standard 15489-1 (2001) and in the draft ISO/PDTS 23081.

Contextual and structural metadata of different processes can be described with same concepts. In the legislative environment, for example, the legislative process for preparing a criminal law can be described to a great extent with the same concepts as the process for preparing a new traffic law. Also the document structures (structural metadata) in the processes are for the major documents the same. Concepts for describing semantic metadata of the documents produced in the processes however differ.

To be able to identify appropriate concepts shared in a community in a business process, the process and its components have to be analyzed and described. Methods for analyzing information management in business processes, particularly for document-centric environments where major portion of the information created in the process activities is recorded in documents have been described in Salminen (2003).

Figure 1 describes the information flow in a business process environment and at the same time a metamodel we have used to gather the concepts related to business processes in an environment. The oval represents the activities of the process, the rectangles three types of information repositories: actors, content items, and systems. An *activity* is a set of actions performed by one or more actors in a process. *Actors* are the performers of activities. An actor is an organization, a person, or a software agent representing a person or an organization in the activities. *Systems* consist of the hardware and software applications used to support the performance of activities.

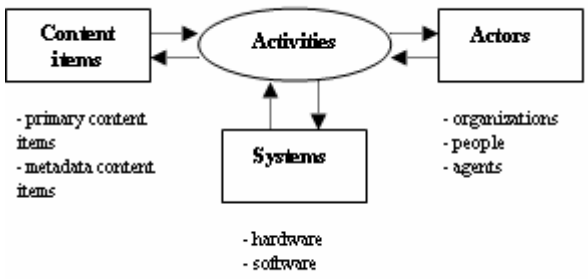

Figure 1: Information flows in the activities of a business process.

In Figure 1 *content items* consist of stored data produced and used in the activities by actors. To be able to handle metadata as resources in the same way as other recorded content items, the metadata should also be recorded as content items. Therefore the content items are divided in the figure into two types, primary content items and metadata content items. Information is depicted by the dashed arrows. Information needed and produced during activities is stored in documents and other content items, in the heads and experience of people, in the organizational culture, and in systems.

Currently we are investigating the possibilities of RDF to be used as the description language for the concepts of a process environment. In the following section we will briefly describe RDF and RDF Schema. Then we will evaluate the opportunities offered by RDF/RDFS and problems related to the definition of RDF schemas and schema concepts.

3 RDF AND RDF SCHEMA

RDF is a model for describing metadata about resources. According to the specification (Manola & Miller, 2004), a *resource* is anything that can be identified on the web. As the mechanism to identify resources RDF uses URI references.

RDF describes metadata as simple (resource, property, value) triples. The triples are called *statements*. The RDF specification defines a graphical representation for RDF descriptions as graphs, and a textual representation using XML syntax.

RDF Schema is a language for defining vocabularies intended for use in RDF statements. RDF Schema vocabulary consists of classes, subclasses and properties, which can be used to define complicated term hierarchies. Properties are defined with range and domain qualifiers. For example, a property *creator* can be defined with a class Document as a domain value and a class Person as a range value.

Below an example of an *RDF process schema* is given. The example is a part of a schema describing the Finnish legislative process. The schema includes classes LegislationProject, LegislativeDocument and ProcessPhase. Properties name, identifier, date, createdInPhase and createdDuringLegislationProject apply to the class LegislativeDocument, as can be seen in the domain definition. Two last properties have range definitions, which indicate to which class the value of the property should belong.

```
<?xml version='1.0' encoding='UTF-8'?><!DOCTYPE
rdf:RDF [<!ENTITY example
"http://www.legislationexample/rdf#">]>
<rdf:RDF
xmlns:rdf="http://www.w3c.org/1999/02/22-rdf-
syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-
schema#">

<rdfs:Class
rdf:about="&example;LegislationProject"/>

<rdfs:Class
rdf:about="&example;LegislativeDocument"/>

<rdfs:Class rdf:about="&example;ProcessPhase"/>

<rdf:Property rdf:about="&example;name">
    <rdfs:domain
rdf:resource="&example;LegislativeDocument"/>
</rdf:Property>

<rdf:Property rdf:about="&example;identifier">
    <rdfs:domain rdf:resource="&example;
   LegislativeDocument"/>
</rdf:Property>

<rdf:Property rdf:about="&example;date">
    <rdfs:domain rdf:resource="&example;
   LegislativeDocument"/>
</rdf:Property>

<rdf:Property
rdf:about="&example;createdInPhase">
    <rdfs:domain rdf:resource="&example;
   LegislativeDocument"/>
    <rdfs:range
   rdf:resource="&example;ProcessPhase"/>
</rdf:Property>

<rdf:Property
rdf:about="&example;createdDuringLegislationPro
ject">
   <rdfs:domain
rdf:resource="&example;LegislativeDocument"/>
   <rdfs:range
   rdf:resource="&example;LegislationProject"/>
</rdf:Property>

</rdf:RDF>
```

Figure 2 demonstrates the design of the process schema using Protégé ontology editor (available at http://protege.stanford.edu/). The left pane shows the RDF classes of the process schema. The process phases have been described by the class ProcessPhase and its four subclasses. In addition to the classes in the textual schema, there are classes for a person and for two organizational actors ministry and committee. In the middle pane there are the instances of the schema (the actual RDF descriptions). The pane on the right shows the properties used to describe a legislative document instance of the class LegislativeDocument.

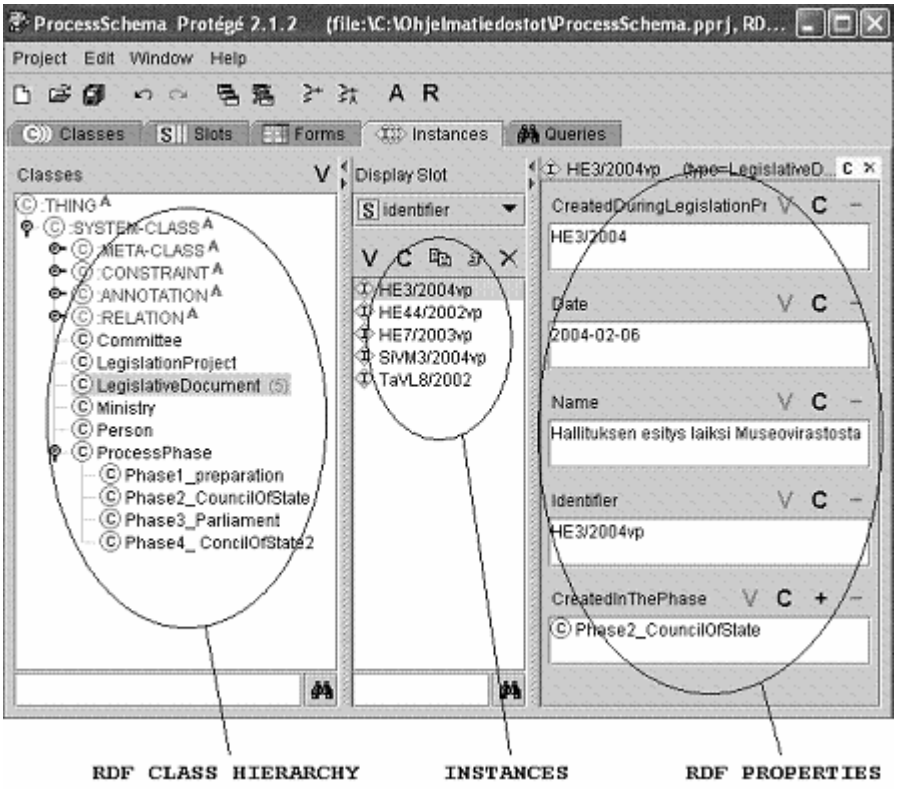

Figure 2: Designing an RDF process schema with Protégé ontology editor.

RDF Schema offers a simple mechanism to define ontologies (Klein, 2002; Volz, Oberle & Studer, 2003). Compared to some other ontology languages, RDF is however limited. The major limitation is the lack of reasoning support, which RDF does not include. To get more capabilities for describing resources, RDF Schema should be developed further or its use should be extended with richer schema languages. According to Manola and Miller (2004) useful additional capabilities would be:

- cardinality constraints,
- adding transitivity definition to a property,
- adding a definition that a property is a unique identifier,
- definition of two classes or instances representing the same class or individuals, respectively,
- definition of constraints on the range or cardinality of a property that depend on the class (e.g. the range of the property identifier differs if the described class is LegislativeDocument or LegislationProject), and
- ability to describe new classes in terms of combinations (e.g. unions and intersections) of other classes, or to define that two classes are disjoint (e.g. no resource is an instance of both LegislativeDocument and LegislationProject classes).

4 OPPORTUNITIES IN RDF

RDF offers many kinds of opportunities for describing metadata for a business process. In this section we evaluate these opportunities.

The **XML syntax** of RDF makes RDF descriptions machine-readable. The XML syntax brings several well-known advantages, such as system independence and possibility to be used as a format in the data interchange of various systems.

RDF together with RDF Schema supports the description of **various types of metadata** including semantic, contextual and structural metadata. Different RDF descriptions can be attached to a single resource. Thus **a collection of resources can be described for different user groups by different terms**. In an inter-organizational environment there can be many groups of actors with varying terms for the same concepts. For example, in a legislative environment citizens and lawyers could have their own schemas.

In complex interorganizational processes the metadata schemas have to be created gradually. RDF supports that kind of proceeding, because RDF schemas can be flexibly **extended** and also **merged** with other vocabularies from various sources with

the help of XML namespaces (Hunter & Lagoze, 2001; Candan, Liu & Suvarna 2001).

The RDF/RDFS has remarkable **flexibility** compared to other modeling approaches. Instead of describing classes with certain properties, RDF Schema introduces properties as first-class entities like classes. Properties apply to specific classes according to domain and range specifications.

As an ontology language RDFS facilitates buiding **intelligent semantic web services**, such as semantic browsing service and semantic recommendation (Hyvönen & al., 2004; Quan and Karger, 2004). Semantic recommendation means that additional related links are shown to the user parallel to the original search result. According to Middleton, Alani, Shadbolt and De Roure (2002), people often feel it difficult to express what they want or what they are looking for. Semantic recommendation might be a useful tool for information retrieval in business process environments.

5 CHALLENGES IN DESIGNING RDF SCHEMAS

Noy and McGuinness (2001) state that probably the most difficult task in the schema or ontology design is to find **consensus** between people in the community. People may have completely different views about important and central terms in the organization and on the domain. In a community of several organizations achieving consensus may be impossible. There are cases where someone should be in a position to make decisions on the schema concepts when representatives of the community do not reach agreement. Identifying **the owner of the process** environment might solve the problem, but getting agreement of the owner is sometimes extremely difficult in an interorganizational process. In the case of the Finnish legislative process no process owner has been identified. The needs for standardization and integration throughout the process have however initiated discussions about the possible owner.

Heterogeneity of users in an organization is also a challenge in schema design. Heterogeneity concerns, for example, the roles of people in work processes and the language they are accustomed to use in communication.

Schema design experts hardly are experts of the target business process. The schema designers have to do very close **cooperation with users** to learn the concepts and needs of the environment. It may however be hard to find from a large set of actors the users having time to collaborate and having enough expertise on the area. If the users have no knowledge about metadata or ontologies, it is sometimes difficult to motivate them to spend their time in the design work.

Ontologies introduced for semantic web usually describe rather static resources like, for example, museum items in the MuseumFinland service (Hyvönen & al., 2004). The items themselves do not change, only the meanings given by people may change. The nature of processes and also the nature of documents created in the processes is **dynamic** and causes thus more frequent update needs for the schema. Also according to the draft ISO/PDTS 23081 **metadata schemas need to be continuously updated** to reflect changes in an organization and in the business.

In many business processes the content of data resources varies so much that the **development of content ontologies** with detailed semantic metadata covering those resources is extremely difficult. The number of concepts easily increases over the limit to be **systematically updated**. In the legislative environment, for example, the legal terminology is huge and the current legal terminology in Finland is closely related to the EU legal terminology. Relations between various laws are continuously changing and totally new areas evolve to be regulated by laws (such as gene technology).

6 IMPLICATIONS

Our analysis has clearly shown that there are interesting opportunities in using RDF to support information management in business processes but the problems for creating the RDF schemas and descriptions are extremely difficult, especially in the cases of complex interorganizational processes. These are however also the cases where the needs for new solutions are most urgent and even minor enrichment of the process environment by well-designed metadata might serve the work in the process. Thus we argue that careful analysis to identify the most essential core metadata would be important in business process environments.

Our experience and analysis has shown the difficulty of developing ontologies to describe the meaning of content items when the content concerns various domains. The number of concepts potentially needed in the ontology may be huge and finding agreement about the standard ontology may be extremely difficult. In ontologies describing the process context of the content items, the number of concepts needed is evidently smaller.

In developing process schemas it is important to look for possible existing standard terminologies. For example, in spite of the many changes taken

place during the last decades in the Finnish legislative work, there still are stable terms for a number of central concepts. The terms are described in a published dictionary (Torniainen, 1999).

Communication between schema designers and people representing the organizations involved is important during the schema design. Views of various organizations should be taken into account. Coordination of the collaboration is essential to facilitate participation of all interested parties.

7 CONCLUSION

In interorganizational business processes a major problem is the heterogeneity of software applications and lack of interoperability. Semantic web technologies offer, at least in principle, a solution to the problem. Building semantic web solutions essentially means developing standardized metadata solutions. In the paper we discussed the metadata types needed in business processes. RDF is the model intended to be used for semantic web metadata. In the paper we analyzed the possibilities to use RDF to support information management in business processes and the challenges in designing schemas for the purpose. In spite of the major problems in the schema design and maintenace of ontologies, the use of RDF seems to offer interesting possiblities. Therefore building and evaluating experimental solutions in real business process environments will be important. Our future work will include designing and evaluating experimental RDF schemas and descriptions for the legislative process.

REFERENCES

Berners-Lee T., Hendler J. & Lassila O. 2001. The Semantic Web. Scientific American 284(5), 34-43. Retrieved July 15, 2004, from http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21

Bray T., Paoli J., Sperberg-McQueen C., Maler E. & Yergeau F. 2004. Extensible Markup Language (XML) 1.0 (Third Edition). W3C Recommendation. Retrieved from
http://www.w3.org/TR/2004/REC-xml-20040204/

Brickley D. & Guha R.V. (Eds.) 2004. RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation Retrieved June 17,2004, from http://www.w3.org/TR/rdf-schema/.

Candan K., Liu H. & Suvarna R. 2001. Resource Description Framework: Metadata and its applications [Electronic version]. *ACM SIGKDD Explorations Newsletter* 3(1), 6-19.

Gruninger M. & Lee J. 2002. Ontology applications and design [Electronic version]. *Communications of the ACM* 45(2), 39-41.

Hunter J. & Lagoze C. 2001. Combining RDF and XML schemas to enhance interoperability between metadata application profiles. *In Proceedings of Tenth International Conference on World Wide Web*, Hong Kong, 457-466.

Hyvönen E., Saarela S., Viljanen K., Mäkelä E., Valo A., Salminen M., Kettula S. & Junnila M. 2004. A portal for publishing museum collections on the Web. *In Proceedings of ECAI/PAIS 2004*, Valencia, Spain (forthcoming).

ISO 15489-1. 2001. International Standard. Information and documentation – Records management. Part 1: General.

ISO/PDTS 23081. 2003. International Standard / Proposed Draft Technical Specification. Information and documentation – Records Management Processes – Metadata Records. Part1: Principles.

Klein M. 2002. Interpreting documents via an RDF Schema ontology. *In Proceedings of the 13th International Workshop on Database and Expert Systems Applications*, Aix-en-Provece, France, 1-5.

Manola F. & Miller E. (Eds.) 2004. RDF Primer. W3C Recommendation. Retrieved from http://www.w3.org/TR/2004/REC-rdf-primer-20040210/

Middleton S., Alani H., Shadbolt N. & De Roure D. 2002. Exploiting synergy between ontologies and recommender systems. *In Proceedings of Semantic Web Workshop 2002 at the eleventh International World Wide Web Conference,* Hawaii, USA.

Noy N. & McGuinness D. 2001. Ontology development 101: A guide to creating your first ontology. Retrieved June 14,2004, from KSL (Knowledge Systems Laboratory) of Stanford University Web site: http://protege.stanford.edu/publications/ontology_development/ontology101.pdf

Quan D. & Karger R. 2004. How to make a Semantic Web browser. *In Proceedings of the 13th Conference on World Wide Web*, New York, USA, 255-265.

Salminen A. 2003. Towards digital government by XML standardization: Methods and Experiences. In Proceedings of the XML Finland 2003, Kuopio, Finland, 5-15.

Torniainen P. (Ed.) 1999. Valtioneuvoston sanasto. Edita. Helsinki.

Volz R., Oberle D. & Studer R. 2003. Views for light-weight Web ontologies. *In Proceedings of the 2003 ACM Symposium on Applied Computing (SAC)*, Melbourne, FL, USA, 1168-1173.