Transparent Knowledge Interchange on the Semantic Web

Martin Svihla, Ivan Jelinek

Abstract: *In spite of a large development of semantic web technologies and a deployment of the first complex applications, the semantic web is still not being widely accepted by the web community and its current potential is not full recognised. One vision of the semantic web is to create a platform for automatic and transparent knowledge sharing. In this paper we discuss a model that uses basic semantic web technologies for knowledge interchange. The architecture of the system is designed to hide all the complexity of the semantic web, so that it can be easily adopted and used in existing web applications. The proposed model and an implemented proof-of-concept application, are suitable for knowledge interchange among web communities that share the same knowledge domain on the web.*

Key words: *Semantic Web, Data Interchange, Software Design.*

INTRODUCTION

The semantic web standards RDF [1] and OWL [13] were deployed as W3C recommendations almost one year ago and there are many available tools that work with these technologies (e.g. Protege[1], Jena[2], 3store[3]). As a result of strong effort of the semantic web community, many complex and useful applications were deployed (CS AKTive Space [12], Finnish Museums on the Web [6], to name a few). Although these applications clearly demonstrate possibilities of the semantic web, the web public doesn't widely use its current potential.

There are various reasons for this situation, we name just a few of them:

- The semantic web is a quite new vision.
- Not all problems in the field are satisfactorily solved.
- The technology is considered to be complex and complicated by people outside the semantic web community.

We reason that even in spite of these facts the principles of the semantic web can be used in real world applications, in our case in the field of knowledge sharing. Our work addresses mostly the last issue from the previous list. We reason that it is still not easy to use semantic web technologies to extend current web applications with semantic features (also in [5]). There are indeed matured tools for a metadata generating, retrieving or storing, but these tools are much more complex than tools for development of classical web applications.

As mentioned in [8], one of tasks for the semantic web researchers is finding a way how to hide a complexity of the semantic web technologies from the users. The web became so wide-spread and popular thanks its simplicity.

In this paper we propose a model that enables automatic knowledge sharing among web portals in one particular knowledge domain. The architecture of our proposal hides all complexity of the semantic web technologies from a web application developer, so that it can be easily integrated into existing web applications. In our proposal it is a responsibility of domain members to generate metadata and to process them on the other side, but we provide the tool, that would simplify the integration of the semantic web technologies for knowledge sharing.

The next section describes motivation for our work in more detail. Then the architecture of our system, data format and implementation details are discussed. Before the conclusion we compare our contribution to related works.

[1] http://protege.stanford.edu/
[2] http://jena.sourceforge.net/
[3] http://sourceforge.net/projects/threestore/

MOTIVATION AND PROBLEM OVERVIEW

There are knowledge about every subject of a human activity on the web. In this information space web sites and portals with the same interest naturally create knowledge domains. Such a domain can consist of electronic journals with one topic, cultural institutions as cinemas and museums or web sites about the same sport.
It is a usual practise that web sites in one domain want to share their knowledge. However, the information published from web sources cannot be automatically interchanged, merged or processed. It is because of data formats (mostly HTML) do not express an information meaning understandable for computer applications.
The semantic web was invented to overcome this drawback [2] and the semantic web technologies can help with an automatic knowledge sharing on the web.
In our work we propose system that enables knowledge sharing in one knowledge domain, where information providers in a particular domain are interested in information interchange. This is the case when for example:

- Web portals with event calendars want to share event information.
- Wholesaler needs to share its product catalog with retailers.
- Electronic museums or galleries want to merge their collections.

As mentioned before, the semantic web technologies are suitable for such tasks. In fact, nowadays they are often used for knowledge sharing.
However, if these technologies should be applied for an information interchange, the web application programmers from all participants of the knowledge interchange in the domain have to understand them. Due to the reasons mentioned in the introduction, this is not a case yet.
The main contribution of our work is the architecture that hides complexity of the semantic web technologies from web application programmers. This architecture preserves the power of RDF and OWL, but a programmer that uses our system doesn't necessarily understand these languages. Based on this idea our system for a knowledge interchange can be easily deployed to existing web sites in the domain.
We also implemented a system for one particular domain - the information about martial art seminars. There are various martial arts and it is common that so called seminars are organized, which workshops are kept by one teacher with a high qualification.
Almost every martial art club has its own website with a list of such events as a date, location, name and qualification of a teacher, schedule and other details. Since these presentations are only in HTML, the event calendars cannot be aggregated and a further processing of these information is not possible.
When we refer to the practical implementation of our model in following text, we refer to the system for a knowledge sharing in the domain of martial art seminars. This system is a software package -- a programming library, which can be easily integrated to existing web presentations of martial art clubs and organizations to enable semantic interchange of seminar calendars. Then the calendars from clubs can be aggregated in the list on an organization website, events from other organizations can be included to that list and all items can be semantically sorted or retrieved.

KNOWLEDGE SHARING MODEL

To enable a transparent semantic knowledge sharing in a particular domain, we provide a software package for communication participants. This programming library, when integrated in existing web application, can generate an information stream, read this stream and process it. Knowledge flowing between publisher and consumer are in RDF format, specified by a domain ontology.
However, all details of this functionality are hidden from a web application developer that

would use this library to enrich the web application with the knowledge sharing capability.
The one and only part of the library visible to this programmer is its API. This API is object based representation of a particular domain. It follows a structure of a modelled reality rather than a structure of an ontology under the library. That means the programmer can use the library and its API intuitively and does not have to understand the semantic web technologies that lay beyond.
The details about this concept are discussed in following sections.

DATA FORMAT AND ONTOLOGIES

Since we work on the knowledge sharing, the data format for knowledge interchange will be RDF [1], the semantic web language developed for a knowledge description.
A vocabulary used in RDF document is specified by an ontology. Finding a common ontology for the entire web is not possible, but when we focus on a particular knowledge domain, we can easily specify terms and rules necessary for knowledge sharing.
Of course, in the decentralised space of web, it is possible that one domain will be described by various ontologies. This problem is being solved by the ontology mapping [7] However, since we build a tool for a knowledge interchange in one domain, we must also define a language of the communication. That also means we need only one ontology and there is no need for the ontology mapping in our case.
The usage of the one and only ontology is very common in today's semantic web applications, this simplification was also used e.g. in CS AKTive Space [12], where the domain was the computer science. One simple ontology is also base of a success of RSS (RDF Site Summary) [9]. As a language for our seminar ontology we chose OWL [13] and we used Protege OWL Plug-in for its editing.
When building the ontology for martial art seminars, we composed it from as many existing ontologies as possible. As a base for our ontology we chose RSS 1.0. This ontology is based on RDF and it has many extension modules. One of these extensions is an event module [10], which perfectly fits to the event calendar of our domain. For describing teachers in the ontology we used some parts of FOAF ontology [4] and all these elements we put together with our own classes and properties.
Another reason for extending RSS is the fact that RSS became very popular format for a syndicating of various lists. That means we can use RSS native tags (as item, title or description) to publish information about seminars in RSS format, so the event list would be accessible also by RSS aggregators, which makes the shared knowledge readable directly for human users.
This added value can be another argument for adapting of our knowledge interchange programming library.
When an ontology is defined for a particular domain, we can build a system for knowledge sharing, which uses this ontology. The architecture of this system follows in the next section.

SYSTEM ARCHITECTURE

Since we are solving the problem of knowledge sharing, there are two types of participants. The one who publishes a knowledge and the one who takes and processes it. We call the former *the publisher* and the latter *the consumer*. Of course, one particular participant can play both roles at the same time in the knowledge interchange, but we will use this classification in the following text.
The publisher publishes its data on the web in RDF format. The consumer interested in this content fetches the RDF document from the particular URL, parses it and processes knowledge.
This simple schema (depicted on figure 1) is very common, it is used for example in RSS syndication.

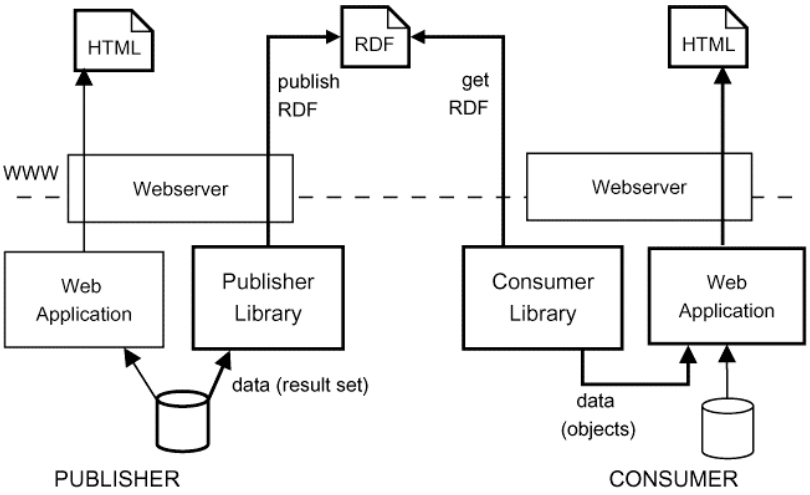

Fig. 1: Knowledge interchange

Even there are two roles in the knowledge sharing; we provide only one programming library to enable the interchange. This library is based on the created domain ontology and it is able to create an RDF document, publish it, fetch it from the web, parse it and process it (a detail architecture of the library is in the figure 2).

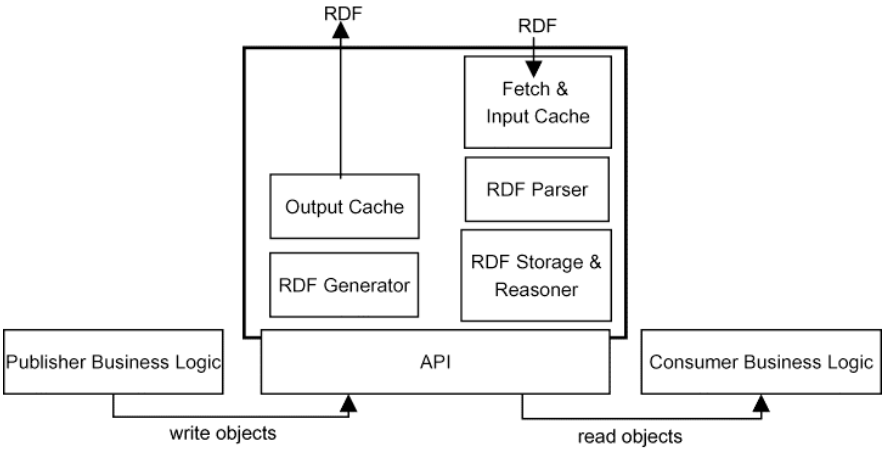

Fig. 2: The library architecture

But an application programmer who uses the library does not work directly with document object model of RDF document. The only part of the library exposed to the programmer is an API, which hides all complexity of the semantic web technologies. This API provides following main features:

- Accessing an RDF document through objects representing knowledge.
- Publishing an RDF document on the web.
- Fetching an RDF document from specified URL.
- Caching an RDF document (on the both sides -- on the side of publisher or consumer).
- Setting programming library specific options – e.g. caching policies or RDF document encoding.

The main part of the API is the interface for accessing knowledge as a collection of objects. The application programmer can build or read RDF document, while manipulating with objects. The object structure is mapped to the structure of an ontology, but this fact is hidden from the application programmer. Moreover, the object hierarchy in API follows the

modelled reality rather than ontology structure, what makes the programming interface more friendly to the programmer.

FLEXIBILITY AND SCALABILITY

Once an API is specified and the library is used as part of web applications over the world, any changes of the API must be considered carefully. In fact only API extension is allowed, because any changes of the interface would damage a code of library users. From this point of view the architecture is quite restricted.

On the other hand, even the API is almost unchangeable, everything under it can vary. For example, the consumer's part of the library can consist of a simple XML parser or a powerful RDF library with an RDF repository and an inference engine.

When the two participants of knowledge sharing incorporate our system, they creates data connection, which is based on the semantic web principles.

The most simple model of such interchange consists of one publisher and one consumer and the knowledge flows only in one direction (see Fig. 3a). We did our first tests in such simple environment, but it is everything but the real world use case.

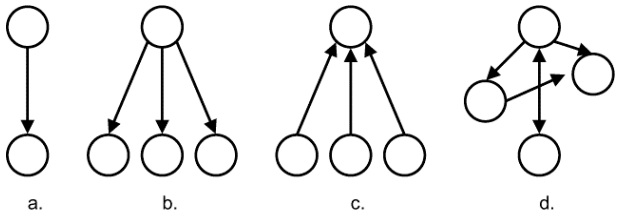

Fig. 3: Knowledge sharing models

However, this model can be extended by more participants and then various application scenarios arise.

A broadcasting scenario with one publisher and many consumers is in the Fig. 3b. Currently we deploy a practical application of our system, when one wholesaler publishes a product catalog for retailers so that they can make this catalog part of their e-shops.

A completely different case is the one with many publishers and one consumer (see Fig. 3c). This is common when information from many small websites are aggregated in one large knowledge base of a domain portal.

And of course, all these classes can be combined, so the mixed model (Fig. 3d) with any possible variations can be built.

When the software for the knowledge sharing is available for a particular domain, it is technically possible to build any of discussed models and it is no problem for new participants to join an existing model later.

CONCLUSION AND FUTURE WORK

The participation of common web sites on the semantic web vision is not a common case yet, even the semantic web standards are successfully deployed and there are many useful semantic web tools available.

We want to contribute to the semantic web growth by a proposal of a model that hides complexity of the semantic web technologies. We provide the programming library that enables a transparent knowledge sharing among web sites from one knowledge domain. In this library the semantic web technologies are hidden under the object oriented API, which can be used more intuitively than a direct interface to the RDF or OWL structures.

We also tried to demonstrate that the development of usable application with semantic web features is not a difficult task today by an example of an implementation the library for the martial art seminars sharing. There is a lot of open-source semantic web tools and libraries that can be used or extended in order to build a brand new application.

To reach the usability of our system we simplified many principles of semantic web. Only one ontology for our knowledge domain is used, so there is no need for ontology mapping. Currently there is no inference mechanism on the side of a consumer; even it is possible to implement some.

As a conclusion we present main principles we had in mind during our work:

- The complexity of the semantic web technologies should be hidden from the application programmer.
- The existing specifications and standards (as RDF, OWL, RSS or FOAF) should be respected.
- The existing tools should be reused for faster a software development.

In the future we want to deploy a framework that would semi-automatically generate the knowledge sharing library based on a given ontology.

Also the RSS aggregator capable to show tags extending a classical RSS structure is being implemented.

ACKNOWLEDGEMENTS

Research described in the paper was financially supported by the FRVŠ under grant no. 1804/2005 and by internal grant of Czech Technical University (IGS) under the external number CTU0507513. It is part of the advanced web technology research in the research group Webing at Czech Technical University in Prague.

REFERENCES

[1] D. Beckett. RDF/XML Syntax Specification (Revised). http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/, February 2004.

[2] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. Scientific American, 284(5), May 2001.

[3] M. Biddulph. A Semantic Web Shoebox - Annotating Photos with RSS and RDF. In XML Europe 2003, London, May 2003.

[4] D. Brickley and L. Miller. FOAF Vocabulary Specification. http://xmlns.com/foaf/0.1/, September 2004.

[5] S. Haustein and J. Pleumann. Easing Participation in the Semantic Web. In Semantic Web Workshop, WWW2002, May 2002.

[6] E. Hyvnen et al. Finnish Museums on the Semantic Web: The user's Perspective on MuseumFinland. In Museums and the Web Conference 2004, Arlington, VA, USA, 2002.

[7] Y. Kalfoglou and M. Schorlemmer. Ontology mapping: the state of the art. The Knowledge Engineering Review, 1(18):1–31, 2003.

[8] M. Lytras. An interview with Eric Miller. SIGSEMIS: Semantic Web and Information Systems, 1(2), July 2004.

[9] RDF Site Summary (RSS) 1.0, Official Specification. http://web.resource.org/rss/1.0/, December 2000.

[10] S. Rough. RDF Site Summary 1.0 Modules: Event. http://web.resource.org/rss/1.0/modules/event/, July 2001.

[11] M. Sabou, D. Oberle, and D. Richards. Enhancing application servers with semantics. In 1st Australian Workshop on Engineering Service-Oriented Systems (AWESOS 2004), pages 7–15. Monash University, Australia, 2004.

[12] M. C. Schraefel, N. R. Shadbolt, N. Gibbins, S. Harris, and H. Glaser. CS AKTive space: representing computer science in the semantic web. In WWW '04: Proceedings of the 13th international conference on World Wide Web, pages 384–392. ACM Press, 2004.

[13] M. K. Smith, C. Welty, and D. L. McGuin-ness. OWL Web Ontology Language Guide. http://www.w3.org/TR/2004/REC-owl-guide-20040210/, February 2004.