INFORMATION ARCHITECTURES FOR SEMANTIC WEB APPLICATIONS

Kimmo Salmenjoki¹, Yaroslav Tsaruk¹, Gurusamy Arumugam²

¹*University of Vaasa, Box 700, 65101 Vaasa, Finland*, ²*Madurai Kamaraj University, Madurai-625021, Tamilnadu, India*

Abstract: In this paper we discuss the issues related to information system technologies. The problems of hardware based services and insufficiency in the widely accepted standards have limited dynamic aspects of industrial applications. Besides the hardware, software and communication considerations one has to look into the system integration, user involvement and service aspects due to the impact of web technologies. We present some of the high level approaches provided by the semantic web technologies in a heterogeneous collaboration network and also discuss the approaches for designing and describing the information architecture of these information systems.

Key words: semantic web; XML; RDF; ontologies; information architectures; software services

1. INTRODUCTION

The innovations in information system (IS) technologies and IT vendor products have been the technological driver in the evolution of industrial applications. The present practices in industrial applications are a colourful collection of different approaches mostly unified by networked TCP/IP only at low level data granularity. To be more service oriented we have to see simultaneously both hardware and software coherently on the organisation of industrial information systems (IIS).

Due to heterogeneous environments the dynamic aspects of industrial applications have been strictly limited. Integration of different data sources and devices has always needed further software development. When different system components need to be integrated either from the business

or process point of view, higher level approaches other than the network and data based integration also have to be addressed. In combining hardware and software, the key issue in information system development has previously been data communication, but the recent penetration of web technologies has shifted the focus towards system integration, user involvement and service and lifecycle aspects. In most cases now-a-days one is eager to consider the life time of industrial product and services which again sets more demand on the abstraction level of software and data for system integration.

The work of Industrial Ontologies Group (IOG) in the University of Jyväskylä, Finland, focuses on establishing industry wide approaches of sharing knowledge and building industrial applications that contain also autonomous and self learning components. Their work on the project Smart Resources has already proved the dire need of new approaches to provide sustainable unifying industrial services for business customers. Related to this wider approach we will describe a case example of applying IOG approaches towards industrial projects containing several actors, see [1, 2].

In what follows, we will go through high level approaches provided by the semantic web technologies in a case scenario taking into account the heterogeneous collaboration network. This paper will address approaches for combining logic and data by using the semantic web based technologies. The case that we present for this approach is figurative, as we describe educational units and their operations from the mutual collaboration point of view, but the approach we suggest should be also valid for other IS development cases.

2. SOLUTION FOR REFINED SOFTWARE AND INFORMATION ARCHITECTURES IN INFORMATION SYSTEMS

2.1 Service based application development in IIS with web technologies

In this chapter we will address the methodologies that have been used for making coherent approaches in both software and data sharing in information systems. Different component approaches like DCOM or CORBA have been successfully used only on the E-business related applications but not on the IIS due to the existence of incoherent industrial environments. This component based approach is only addressing the heterogeneous software dilemma leaving many of the data related dependencies to the very low level in the network.

The developments in web environment is also penetrating the IS applications either via XML features provided by the database vendors or through other higher level data description. More or less in the near future IS development can be seen from two different approaches: either consisting of unified services or shared data. Again the already existing heterogeneous legacy approaches will be an obstacle for the wider applicability of web services to produce more unified shared logic among the individual software industrial application systems. At least this has been the case for the last ten years, though the components have gained wide acceptability in the enterprise and IT vendor communities.

The implicit usage of XML makes the technical exchange of data more transparent. Partly it also makes data visible beyond application boundaries. The wider usage of web services as a basis of software development still requires deeper consideration of software architecture, see [3].

The previously mentioned components approaches have been unified to a new approach of building software with web service components that have been standardized by W3C. This means, in practice, XML and its technologies have become the de facto approach for sharing data both within and between applications. Again on the enterprise side all IT vendors are strongly pushing and promoting the usage of web services as a methodological approach to develop information systems and improve their integration properties in the future. The conservativeness of industrial application will slow these developments in IS. However the thorough penetration of IT in all modern systems implies these developments also in the industrial side.

These crossovers also introduce some problems on the overall architecture and management of system components as possible services with accurately described components and data interfaces. The previously described software architecture does not necessarily account for all the practical implications of the complex applications as distributed applications, see [4].

This discussion has further lead into service oriented architectures and its future application in software factory type manner, see [5]. This approach fits well with the previously described service approach for IS development although it, at the same time, highly complicates the manageability, security and interoperability issues by making the software components again more granular.

2.2 Applying semantic web approach for software services

In the classical web environment the semantic web approach has been mostly developed to contain shared information architecture within some restricted application domains, most notably the library information systems and their organisation. After dozens of years of development in both information description and IT development, especially information data bases and encoding, the global community of library developers have come up, and widely implanted, the usage of Dublin core (DC) as the shared abstract data description in the libraries all over the world. At the present, however, the web technology improvements will question the practical expandability of the Dublin Core to other application domains. For any application domain of semantic web this process should be a reminder of the complexity in developing widely shared vocabularies and their sharing.

Figure 1. Four different approaches for unified service based IS development

In Figure 1 the three different technological approaches for building service based IS are given by the three named arrows. The technical specifications and explanations for the respective standards and efforts are given in [6, 7]. The fourth, shared ontologies approach, is represented by the cloud that contains all the technological approaches used in a coherent way. The introduction of semantic web eases the software complexity, at least in the enterprise and web scenarios, and also improves the previously discussed

low level granularity of the data layer. After the presently on going standardisation process and its acceptance both by the IT vendors and the industrial developers the necessity of unifying software and data structures is the simultaneous application of semantic web and web services. When the IT software development tools become XML and web services enabled the next higher approaches for information architecture will be based on the RDF, RDFS and domain space specific ontologies.

Using RDF can also be seen as a way of building self-aware and proactive data. This is the need put forward by the two previously described views of IS software consisting either from unified services or shared data. Assuming that the semantic web environment tools are available, an essential part of the complexities of both software and information architecture could be addressed in the XML and RDF description of processes as services. In general, from the services and structural point of view the vendor specific versions are the preferred ones, whereas from the operational and standardization point of view the low level granularity approach prevails. In a unified development environment business processes can also be shared in integrated manner. From the previously described web services approach the methodological interest in new standardisation has moved towards distributed computing in heterogeneous environments, see [8]. As an example of the unification that is presently happening, we consider next an example case of the educational domain and it's information architecture. Addressing both software development (Section 2.1) and semantic web aspects (Section 2.2) we address the decentralized software architecture dilemmas.

3. APPROACH FOR INFORMATION DESIGN WITH SEMANTIC WEB-AN EXAMPLE

3.1 MODE approach for analyzing Baltic Sea Network BSN project data

In the University of Vaasa, we are running the project MODE, which addresses Management Of Distributed Expertise. In this project we have analyzed different cases, where several networked organisations share interest and knowledge on common projects. Although educational units collaborate continuously, there are many problems to establish common terminologies among the universities or while working in specific projects as all things are heavily language, culture and practical operational habits.

To simplify the handling, we next introduce the Baltic Sea Network BSN as a case project of MODE and discuss its information architecture in detail. The main purpose of BSN is to combine efforts in sharing education and research operations and interests among the partner universities. The network promotes international co-operation focusing on the following areas: Welfare, Business Skills and Management, Tourism, and Information and Communication Technology always taking into account the sustainable development. To the BSN network belongs about 40 educational organisations. Each of these organisations has a list of courses, which could be available to any of student of BSN institution. We use BSN, as a case project of MODE, to create globally sharable structure data schema for the communication of data among participating organizations. The general project related information architecture will be left out of this paper.

Using the general methodologies in the Figure 2 we will next give a design for an information architecture that encapsulates the BSN project metadata in the educational application domain. As the general approach of Figure 1 is still somewhat unclear, we give here an example of the educational domain shared ontology development process.

Figure 2. Options for metadata placement in information architecture of the educational domain

As described before for generic IS and software development in previous chapter, here we can use web standard or semantic web approaches. In the educational domain the three arrows in the middle of Figure 1 are cut down into two, as the standardizing organizations self contain education suppliers i.e. universities. Like before, also here we have to consider the granularity and access of the specific metadata items. As we see from the Figure 2, in the educational domain, the semantic web has not yet gained any popularity, although web pages and tools have been widely used for delivering information, personal and group based communication in education.

Here we can say that as we are using the standard semantic web approach, all the standards and tools of semantic web will be available to produce the knowledge aware applications for the educational domain. The suggested information design has to correspond to the needs of the semantic web applications that will be developed for the project and educational domain in Figure 2. Besides the classical search engine type applications we will also produce web based information gathering applications that are dynamically linked to the respective home pages of the partner universities of Baltic Sea Network. The semantic web based applications enable intensifying collaboration in specific subject areas. On the research side, semantic web application will be built to link the suggested project proposals (with its tenders) to partners working in the same interest area. Internally each partner can also use this information for internal resource planning. For the information design purposes we will next give more technical approaches to design and describe the BSN project metadata.

3.2 Technical description of the information contained in the BSN

At present the BSN has a website and the Optima learning environment is used for communicating data and information exchange among the partners. Based on the approaches in Figure 2, we will describe the generic educational organization data structure next. The overall information structure is given in Figure 3.

The whole structure presents information about educational units, personnel and courses. The information about an educational unit is presented by classes such as “Organisation” and “Department”. The information about a staff in the organisation is given by the classes “Person”, “Staff”, “Degree”. Class “Person” contains basic information about any one working in the university. The “Degree” class includes data about educational qualifications of a staff working in the university by the attribute

“state” in that class. The class “Staff” relates the person and degree information.

Figure 3. The structure of information contained in BSN project

The structure presented in Figure 3 is next communicated within the network and refined further against the higher level community standards of Figure 2. We next present, in Example listings 1 and 2, the metadata structures related to XML and RDF descriptions for some items of Figure 3.

```

<?xml version = "1.0" encoding = "UTF-8"?>
<!-- Namespace Declarations in XMLSchema -->
<xsd:schema xmlns:xsd = "http://www.w3.org/2001/XMLSchema"
  xmlns:edu = "http://people.jyu.fi/~yatsaruk" version = "1.0">

```



```

<xsd:complexType name = "Organisation">
  <xsd:sequence>
    <xsd:element name = "name" type = "xsd:string" />
    <xsd:element name = "type" >
      <xsd:simpleType>
        <xsd:restriction base = "xsd:string">
          <xsd:enumeration value = "University"/>
          <xsd:enumeration value = "Polytechnic"/>
          <xsd:enumeration value = "College"/>
        </xsd:restriction>
      </xsd:simpleType>
    </xsd:element>
    <xsd:element name="Departments" />
    <xsd:sequence>
      <xsd:element name = "Domain" type = "xsd:string" />
      <xsd:element ref="edu:Location" />
      <xsd:element ref="edu:Staff" />
    </xsd:sequence>
  </xsd:complexType>

<xsd:complexType name = "Department">
  <xsd:sequence>
    <xsd:element name = "Name" type = "xsd:string" />
    <xsd:element name = "Domain" type = "xsd:string" />
    <xsd:element ref="edu:Location" />
    <xsd:element name="Courses" />
    <xsd:sequence>
      <xsd:element ref="edu:Course" />
    </xsd:sequence>
  </xsd:sequence>
</xsd:complexType>
</xsd:schema>

```

} Organisation
definition

} Department
definition

Example 1. Metadata presented in XML Schema format

Description of Organisation and Department is given in Example 1. Both these units are presented as complexType elements. Each of these complexType elements includes a set of elements. The element could be of a basic type or another complex element. The “Type” property is declared as enumerated one. The field location is defined by reference to complex element Location, situated in “edu” namespace. The “Department” unit

contains element such as “Courses”, which describes the list of courses proposed by this department. The element “Courses” is a set of reference to element “Course”, which is situated in “edu” namespace.

On specific application domains the IT vendors will be gradually improving their information granularity in a manner similar to what Microsoft has done on the IT applications and their end users side. In the simple office and enterprise scenario some vendors like Microsoft have provided examples of unified approaches like the .NET Framework as a distributed computing environment or unifications of information either by application to application data sharing or usage of XML namespace based information exchange like sharing data between Microsoft Office applications via the Microsoft Office Namespace Schema. An example domain, where this has been widely used, is the Danish national effort to unify communal system and processes developments by sharing the MS-Office namespaces between different communal actors and applications, see [9]. In that case this makes the MS-Office namespace as the universal information sharing architecture among any information exchange partners. This is the lowest level of creating information architectures which is based notably on shared XML namespace usage.

```
<rdfs:Class rdf:ID="Staff">
  <rdfs:comment>Staff Class</rdfs:comment>
  <rdfs:subClassOf rdf:resource="http://people.jyu.fi/~yatsaruk/world#Person"/>
</rdfs:Class>
<rdfs:Class rdf:ID="Organisation">
  <rdfs:comment>Class for general organization units</rdfs:comment>
</rdfs:Class>
<rdfs:Class rdf:ID="EducationalUnit">
  <rdfs:comment>Class for educational organization units</rdfs:comment>
  <rdfs:subClassOf rdf:resource="#Organisation"/>
</rdfs:Class>
<rdfs:Class rdf:ID="University">
  <rdfs:comment>Class for universities</rdfs:comment>
  <rdfs:subClassOf rdf:resource="#EducationalUnit"/>
</rdfs:Class>
<rdfs:Class rdf:ID="Polytechnic">
  <rdfs:comment>Class for polytechnics</rdfs:comment>
  <rdfs:subClassOf rdf:resource="#EducationalUnit"/>
</rdfs:Class>
<rdfs:Class rdf:ID="College">
  <rdfs:comment>Class for colleges</rdfs:comment>
  <rdfs:subClassOf rdf:resource="#EducationalUnit"/>
```

```

</rdfs:Class>
<rdfs:Class rdf:ID="Department">
  <rdfs:comment>Class for departments</rdfs:comment>
  <rdfs:subClassOf rdf:resource="#EducationalUnit"/>
</rdfs:Class>
<rdf:Property rdf:ID="partOf">
  <rdfs:comment>Part-of relationship</rdfs:comment>
  <rdfs:domain rdf:resource="#Organisation"/>
  <rdfs:range rdf:resource="#Organisation"/>
</rdf:Property>
<rdf:Property rdf:ID="name">
  <rdfs:comment>The name of organisation or department</rdfs:comment>
  <rdfs:domain rdf:resource="#Organisation"/>
</rdf:Property>
<rdf:Property rdf:ID="Location">
  <rdfs:comment>Location of organisation</rdfs:comment>
  <rdfs:domain rdf:resource="#Organisation"/>
</rdf:Property>
<rdf:Property rdf:ID="ContactPerson">
  <rdfs:comment>Contact person of organisation</rdfs:comment>
  <rdfs:domain rdf:resource="#Organisation"/>
  <rdfs:range rdf:resource="#Staff"/>
</rdf:Property>
</rdf:RDF>

```

Example 2. Presentation of metadata structure in RDFS format

The information in RDF format is presented by two elements: classes and properties. The definition of classes presented in the beginning of RDF document. The “Organisation” class is inherited from “Resource”. The parents for class “Staff” is “Person”. The description of properties depicted in the rest of RDF document. The property type Alt is used for describing container where just one option should be selected of the values attributed to the element in the RDF.

All information exchanges that take place in the previously described BSN case project can be seen from two different views: internally from the organization or externally from the BSN project point of view. In internal is a data presentation format used inside each institution, and this is reflected on the web pages of the respective universities. These sets of data characteristics for each educational organisation are specified in their own way. After structuring and presenting this metadata in RDF format it becomes external format of the BSN data. This data and its metadata will be used to enhance the information sharing between the partners and further to describe knowledge based applications.

3.3 Comparing XML and RDF approaches for BSN information architecture

Next we will clarify in detail the advantages and disadvantages of using the XML and RDF based approaches in Example 1 and Example 2.

The examples discussed before show the different presentation formats of the same information. The structuring and presentation of information are different in both formats but namespaces have been used in both of them. In XML format namespaces need not point to anything in the XML Namespace specification. In RDF, the namespace URI reference also identifies the location of the RDF schema. RDF format presents object oriented paradigm. Resource is the top level class. Latest revisions to the RDF specifications allow cycles in class hierarchy which was not there earlier. In XML the information is presented by elements of certain Type. The type can be of two types: simple and complex. Complex contains a set of elements inside. The class is defined by elements and their properties could be another element. It has no defined semantics. Inheritance can not be realized in XML format. However, types can be “extended” or “restricted”, thus defining subTypes. But in RDF along with object and classes inheritance can also be realized. A class can be a subclass of other classes (multiple inheritance is allowed). The inheritance is related to Property. Properties can be subPropertyOf other properties. The type of data used in XML and RDF formats is different. In RDF the core RDF Schema includes “Literals” which is the set of all strings. The latest RDF specification is expected to include XML Schema data types. The data types supported by XML Schema are mainly variations of numerical, temporal and string data types. The XML format allows describing the enumeration of properties using <enumeration> tag. The RDF doesn’t allow such possibility, see [10].

Finally the essential question is on the users of the information architectures and also on the necessary applications that would be using this architecture. In case of the learning systems the evolution of the systems is presently covering the XML as medium of sharable information. Also here the IT vendor supported technologies and tools, most notably the web services, are the likely interfaces that will give access to the granularly refined learning objects, which are shared by the learning communities. Again also here, with the simultaneous usage of semantic web to provide meaning and web services to provide the access, we are able to unify the learning objects into knowledge based educational applications.

4. CONCLUSIONS

We have seen in this paper how the information design is a necessity for building knowledge based applications. Once the information architecture is given then the general methodologies, technologies and related tools can easily enable knowledge based applications in the domain scope. When refining the project metadata in general we will split the metadata into different sections like strategic tasks, human resources and contextual connections. Also here the ontological approach will be combined to the work of Mikko Laukkanen and Heikki Helin, who have built semantic web applications for finding an expert within an organisation [11]. When building the higher part of the ontologies the European Union wide curricula and degree structures would provide models for the sharable ontologies.

As next step in our approach, we will refine the information extraction phase in the Figure 1, so that we can automatically harvest as much of the above data, related to both the partners and their educational offering. We will consider the semantic web software and application needs of this case in more detail in the follow up papers [12, 13].

REFERENCES

1. Terziyan V., Semantic Web Services for Smart Devices in a Global Understanding Environment in: *On the Move to Meaningful Internet Systems 2003*: edited by Meersman R. and Tari Z. (OTM 2003 workshops, Lecture notes in Computer Science, Vol.2889, Springer-Verlag, 2003, pp.279-291)
2. Kaykova O., Khriyenko O., Kovtun D., Naumenko A., Terziyan V., Zharko A., General Adaptation Framework: Enabling Interoperability for Industrial Web Resources, *International Journal on Semantic Web and Information Systems*, Idea Group, ISSN:1552-6283, Vol.1 No.3, July-September 2005 pp.30-62 (to appear)
3. Alesso H.P., Smith C.F: *The Intelligent Wireless Web*, Addison Wesley, Pearson Education, 2002.
4. Singh M.P, Huhns M.V., *Service Oriented Computing: Semantics, Processes, Agents*, Wiley, 2004.
5. Greenfield J., Short K.: *Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools*, Wiley, 2004.
6. World Wide Web Consortium, Semantic Web,
<http://www.w3.org/2001/sw/>

7. World Wide Web Consortium, Web Services,
<http://www.w3.org/2002/ws/>
8. Da Conta M.C., Obrst L. J., Smith K.T.: *The Semantic Web: A guide to the Future of XML, Web Services and Knowledge Management*, Wiley 2003
9. Danish e-Governance Offentlig Information Online,
<http://www.oio.dk/English>
10. Gil Y. and Ratnakar V. "A Comparison of (Semantic) Markup Languages" USC Information Sciences Institute and Computer Science Department; <http://trellis.semanticweb.org/expect/web/-semanticweb/paper.pdf>
11. Laukkanen M. Helin H., Competence manager- applying semantic web in practice in: *Third International Semantic Web Conference (ISWC2004), Proceedings Series: Lecture Notes in Computer Science, Vol. 3298*, edited by McIlraith S., Plexousakis D., Harmelen F. van; <http://iswc2004.semanticweb.org/demos/22/paper.pdf>
12. Salmenjoki K., Terziyan V., Approaching Web Based Information With Semantic Web, Unpublished Manuscript
13. Tsaruk Y., Salmenjoki K., Project Casing Pilot System., Unpublished Manuscript