Zbornik radova 49. Konferencije za ETRAN, Budva, 5-10. juna 2005, tom III
Proc. 49th ETRAN Conference, Budva, June 5-10, 2005, Vol. III

ONTOLOGY-MEDIATED INTEGRATION OF KNOWLEDGE RESOURCES: THE CASE OF MIHAJLO PUPIN INSTITUTE

Valentina Janev, *Mihajlo Pupin Institute, Volgina 15, 11060, Belgrade.*
Sanja Vraneš, *Mihajlo Pupin Institute, Volgina 15, 11060, Belgrade.*

Abstract - *The objective of this research is to design a business integration model suitable for the R&D organizations in high technology industry. This paper proposes a system integration framework for knowledge management at the Mihajlo Pupin Institute. The proposed framework is based on the latest semantic technologies: ontologies and Web Services. The paper introduces the W3C standards RDF/RDFS and OWL and presents the developed ontology as a knowledge model for the Mihajlo Pupin Institute. The ontology is built using open source tool Protégé.*

1. INTRODUCTION

The most important activities in R&D process are governance of research efforts towards innovation (Suh et all, 2004) and effective management of scientific and technical knowledge transfer towards industry (Rodrígues et al., 2004). In order to improve the creativity and speed up product development, research work typically requires internal integration of heterogeneous knowledge resources such as publication database, project management database, human resources database, etc. Furthermore, technology transfer to other research groups or to industry depends on business integration of the partnering companies i.e. the compatibility of their information infrastructures. Therefore business integration is of high importance for R&D organization.

Motivated by the need to build a business integration framework suitable for internal activities of an R&D organization in high technology industry we are designing a mediating layer that will connect knowledge workers in the Mihajlo Pupin Institute (MPI) with the available internal knowledge artifacts. The aim of this paper is to present the tools and the results of business integration framework design. The design is based on the latest semantic technologies i.e. ontologies[1] and Web services[2]. The paper is organized as follows. Section 2 discusses the needs for integration of knowledge resources in R&D organizations. Section 3 introduces the W3C standard Semantic Web languages RDF/DRFS and OWL. Section 4 presents the results of ontology building for MPI business integration framework.

2. INTEGRATION OF KNOWLEDGE RESOURCES IN R&D ORGANIZATIONS

The business process of an R&D company can be characterized as highly knowledge intensive one where existing knowledge is used to facilitate innovation (Suh et all, 2004). Knowledge resources that form the research and development process in a high technology R&D organization include knowledge workers (researchers), knowledge artifacts (software products, prototypes and methods that are results of prior innovation activities of researchers, as well as the accompanying project documentation) and standardized business processes. In order to explain the need for business integration, we will look closely at the innovation process in an R&D organization (Figure 1).

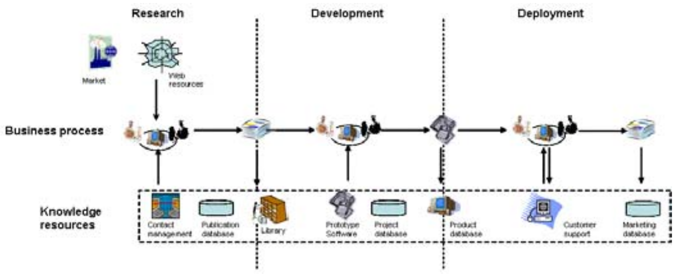

Figure 1. The innovation process.

Innovation process of a high technology product could be divided into three phases: research phase, development phase, and deployment phase.

Research phase could be described as a period of collecting information about the existing solutions for a particular business problem and search for an improved solution of the problem. During this phase researchers consult the internal knowledge bases, as well as the shareware web resources with the aim to become familiar with the problem domain, conduct market research, gain understanding of the existing solutions and propose an innovated solution. The results of this phase could be published or kept strictly confidential in paper or electronic form in publication and project database.

Development phase is a period of implementation of a prototyped solution for a particular business problem in an experimental environment. In this phase software/hardware engineers consulting the internal knowledge bases for existing technical documentation or reusable knowledge objects (hardware, software) and developing new knowledge items (e.g. methods, program interfaces) build a new solution to the problem. Project documentation that accompanies this phase is company private property and for internal use only.

---

[1] Ontology is an explicit formal specification of terms in a domain and the relationships among them (Gruber, 1993).

[2] Web-Services are loosely coupled reusable software components that semantically encapsulate discrete functionality and are distributed and programmatically accessible over standard internet protocols (W3C 03).

Deployment phase is putting the prototyped solution into practice. Here the knowledge or (software/hardware) product is exported to the client along with product documentation that will help clients to easily apply the tool in their working process. Software engineers are engaged in knowledge transfer process until the users on client side are capable of carrying out the activities independently. The experience gained in the knowledge transfer process is documented as a case study and shared with other engineers for future use.

Analyzing the innovation process, generic activities could be identified as searching for knowledge items (human expertise, program code, etc), publishing the knowledge items in appropriate knowledge base so that it could be accessed later, retrieving the knowledge items, accessing the e-learning system, etc. These activities can be implemented as intranet Web services that play the role of programmable application interfaces to existing systems (ERP, document management system, e-learning system, business contact management system, etc) accessible using standard Internet protocols. They will be dynamically invoked by different business tasks thus providing support for automating marketing, sales, customer support, research, and development process. In order to be recognized as useful and compatible for a specific task, they have to be described with terms common to business domain. Hence the first task in business integration framework design is building a knowledge model i.e. ontology for the R&D organization.

3. SEMANTIC TECHNOLOGIES FOR BUSINESS INTEGRATION

The term *business integration* usually refers to plans, methods, and tools aimed at modernizing, consolidating, and coordinating the computer applications in an enterprise (WhatIf definition) e.g. migrating the existing legacy applications and databases in order to be used in Semantic Web. Business process integration relies on suites of software with the capability to define and map processes, develop and deploy processes, and execute, administer, monitor and report. Technologies used for business process integration include Web servers, application servers, business rule engines and emerging Semantic Web technologies. Semantic Web is a set of standards and tools which seek to improve the semantic awareness of interconnected computers by using a systematic representation of knowledge and attributes about things that exist. This representation or model is referred to as ontology. Current state-of-the-art business integration services are based on open standards (Java/J2E, XML, XSLT, WSDL) as well as on semantic technologies (Omelayenko, 2002; Tsai et all, 2003; Turner et all, 2003). Semantic technologies include RDF/RDFS, DAML+OIL, WSDL, OWL, etc.

In February 2004 The World Wide Web Consortium (W3C, 2004) announced final approval of two key Semantic Web technologies, the revised Resource Description Framework (RDF) and the Web Ontology Language (OWL)[3]. RDF is a general-purpose language for representing information on the Web. Information is described in terms of objects ("resources") and relations between them using RDF Schema. RDF Schema servers as the meta language or vocabulary to define properties and classes of RDF resources. RDF Schema uses Uniform Resource Identifier (URI) references for naming. An URI reference is a string that represents a URI, i.e., name or address of an abstract or physical resource on the Web. A URI reference may take the form of a full URI, e.g.

```
rdf:resource=http://www.w3.org/2000/01/rdf-schema#
```

or a portion of that resource e.g.

```
rdf:resource="http://www.w3.org/2000/01/rdf-
             schema#Resource"/>.
```

Thus a general concept "Person" in RDF Schema is described as follows.

```
<!DOCTYPE rdf:RDF [
    <!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-
syntax-ns#'>
    <!ENTITY rdf_
'http://protege.stanford.edu/rdf'>
    <!ENTITY rdfs 'http://www.w3.org/2000/01/rdf-
schema#'>
]>

<rdfs:Class rdf:about="&rdf_;Person"
    rdfs:label="Person">
   <rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
</rdfs:Class>
```

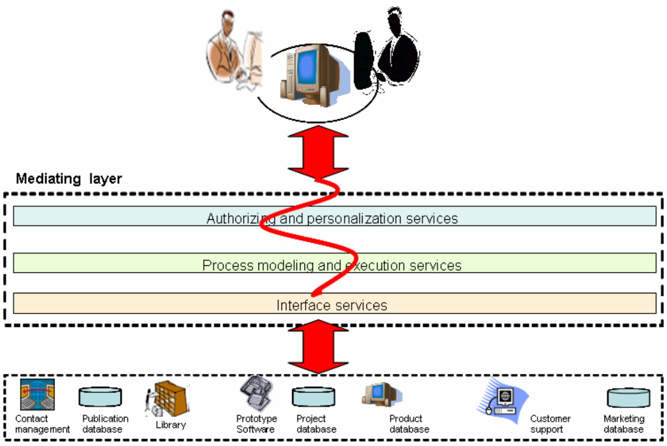

Figure 2. The mediating layer.

OWL extends RDF by adding more vocabulary for describing properties and classes such as relations between classes (e.g. disjointness), cardinality (e.g. "exactly one"), equality, characteristics of properties (e.g. inverse relation), etc. Thus, to point out that one relation is inverse to another the following OWL syntax will be used:

```
<owl:ObjectProperty rdf:ID="presents">
    <rdfs:range rdf:resource="#ServiceProfile"/>
```

[3] RDF and OWL are Semantic Web standards that provide a framework for asset management, enterprise integration and the sharing and reuse of data on the Web. These standard formats for data sharing span application, enterprise, and community boundaries, since different types of users can share the same information even if they don't share the same software (W3C, 2004).

```
    <rdfs:domain rdf:resource="#Service"/>
    <owl:inverseOf>
      <owl:ObjectProperty rdf:ID="presentedBy"/>
    </owl:inverseOf>
  </owl:ObjectProperty>

 <owl:ObjectProperty rdf:about="#presentedBy">
    <rdfs:domain rdf:resource="#ServiceProfile"/>
    <rdfs:range rdf:resource="#Service"/>
    <owl:inverseOf rdf:resource="#presents"/>
  </owl:ObjectProperty>
```

4. ONTOLOGY-MEDIATED INTEGRATION FRAMEWORK FOR MIHAJLO PUPIN INSTITUTE

The mediating layer that aims to connect the knowledge workers at the Mihajlo Pupin Institute with knowledge objects in a uniform way should consist of business rules, logic, processes, and knowledge models that are common to the software application used in R&D process, however, unlocked from any particular application. Decoupling the knowledge from software application will enable design of generic services that can be reused without regard to the type of business content passing through them. In general, the mediating layer (see Figure 2) will be composed of:

- Authorizing and personalization services,
- Process modeling and execution services,
- Interface services.

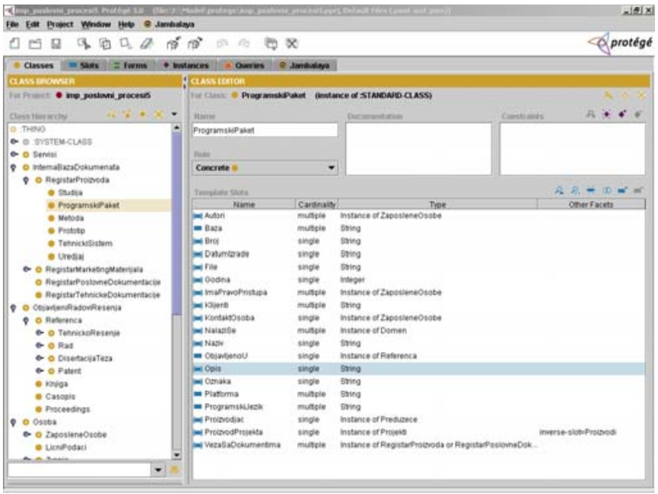

Figure 3. MPI classes presented with Class browser in Protégé 3.0.

Services will use the same ontologies, i.e., knowledge models that represent agreed domain semantics. Roughly, ontologies correspond to generalized database schemes made up of concepts, and the relations between them, that are relevant for a specific domain of knowledge. However, ontologies enable presentation of the structure and meaning of much more complex objects than common databases and are therefore well-suited for describing heterogeneous, distributed and semi structured information sources such as knowledge artifacts of an R&D organization. The presentation of the developed MPI ontology follows.

The MPI ontology is developed using Protégé 3.0. Protégé (Protégé, 2005) is an open-source, Java tool that provides an extensible architecture for the creation of customized knowledge-based applications. In Protégé concepts are named classes and they are described with properties. Classes are either abstract or concrete linked to istances. Properties (Figure 3) are drawn from atribute base that is common to all clases. Attributes are values of type Boolean, Float, Integer, String, Symbol or relation to other clases or istances of clases. Relations could be graphically observed via Jambalaya add-in for Protégé (Figure 4).

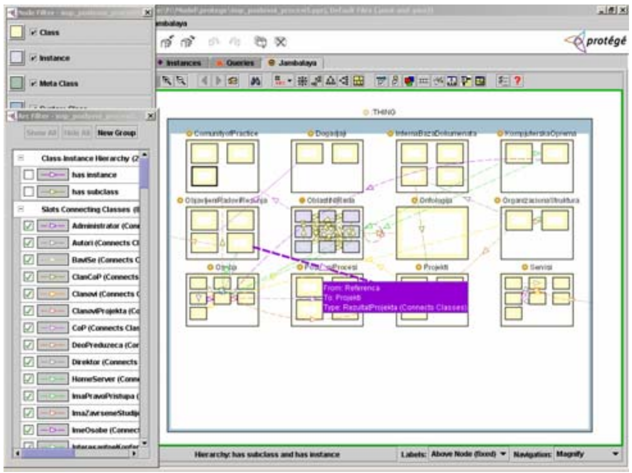

Figure 4. Graphical presentation of the relationships among MPI classes presented with Jambalaya add-in for Protégé.

The MPI ontology is designed having in mind the nature of processes in R&D organizations in high tech industry. Among the main concepts, we point out the following abstract concepts: Product database, Document database, Person, Reference, Community of Practice, Research Area, Events, Project, Business Process, Semantic Web Service, Computer Domain, and Organizational Structure. Abstract classes are super classes of concrete classes that represent concrete objects, terms, individuals or group of individuals. Concrete classes are defined taking into consideration the specifics of the research organizations, e.g., reporting to Ministry of Science or the adopted ISO standard. Parts of MPI ontology in OWL are given below.

```
<rdf:RDF

xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns="http://www.owl-ontologies.com/unnamed.owl#"
  xml:base="http://www.owl-ontologies.com/unnamed.owl">.
```

Class definition looks like

```
 <owl:Class rdf:ID="Referenca">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="ObjavljeniRadoviResenja"/>
    </rdfs:subClassOf>
    <protege:abstract>true</protege:abstract>
  </owl:Class>
  <owl:Class rdf:about="#Rad">
```

```
    <protege:abstract>true</protege:abstract>
    <rdfs:subClassOf rdf:resource="#Referenca"/>
  </owl:Class>.
```

An attribute property definition id written as

```
<owl:DatatypeProperty rdf:ID="Naziv">
    <rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#stri
ng"/>
    <rdfs:domain>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
           <owl:Class
rdf:about="#OrganizacionaStruktura"/>
          <owl:Class rdf:about="#Projekti"/>
          <owl:Class rdf:about="#OblastiNIRada"/>
          <owl:Class
rdf:about="#InternaBazaDokumenata"/>
          <owl:Class rdf:about="#Ontologija"/>
          <owl:Class rdf:about="#Zvanje"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
<rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#Functio
nalProperty"/>
  </owl:DatatypeProperty>
```

An instance is defined as following.

```
  <DoktorskaTeza rdf:ID="imp_Instance_10077">
    <Koeficijent
rdf:datatype="http://www.w3.org/2001/XMLSchema#stri
ng"
    >R81</Koeficijent>
    <Bod
rdf:datatype="http://www.w3.org/2001/XMLSchema#floa
t"
    >6.0</Bod>
    <Godina
rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
    >1999</Godina>
    <Autori rdf:resource="#imp_Instance_10078"/>
  </DoktorskaTeza>
</rdf:RDF>
```

5. CONCLUSIONS AND FUTURE WORK

Research and development process is highly knowledge intensive. In everyday work knowledge workers have to consult diverse knowledge sources including public Web resources, internal publication and project databases, product database, etc. Hence, an integrated knowledge management framework is a prerequisite for creativity and innovation.

In this paper we have presented the initial result of system integration framework design for the Mihajlo Pupin Institute. The proposed framework is based on the latest semantic technologies: ontologies and Web Services. An ontology as a common knowledge model for all knowledge resources in MPI has been developed using Protégé open source tool. The future work will include design of MPI business processes in a machine understandable language (OWL) and implementation of semantic Web services that will connect the MPI knowledge workers with knowledge objects (software application, documentation) in a uniform way.

REFERENCES

[1] W. Suh, J.H. Derich Sohn, J. Y. Kwak, "Knowledge management as enabling R&D innovation in high tech industry: the case of SAIT", *Journal of knowledge management*, Vol. 8, No. 6, pp. 5-15, 2004.

[2] J.L. Rodrígues, A.R. Castellanos, S.Y. Ranguelov, "Knowledge management analysis of the Research & Development & Transference Process at HEROs: a Public University Case", In *Proc. I-KNOW'04 International conference on knowledge management*, K Tochtermann and H Maurer (eds), pp. 61-67, Graz: J.UCS, Know-Center, (2004).

[3] T.R. Gruber, "A translation approach to portable ontology specification", *Knowledge acquisition*, Vol. 5, No. 2, pp. 199-220, (1993).

[4] W3C- The World Wide Web Consortium, "Web Services Activity", http://www.w3.org/2002/ws/, access: 15 March 2005.

[5] Whatis?com – The leading IT encyclopedia and learning center, http://whatis.techtarget.com, access: 15 March 2005.

[6] B. Omelayenko, "Ontology-mediated business integration", In: *Proceedings of the 13-th EKAW 2002 Conference*, Siguenza, Spain, October 1-4, LNAI 2473, 2002, pp. 264-269. Springer-Verlag .

[7] T. Tsai et all., "Ontology-Mediated Integration of Intranet Web Services" *IEEE Computer*, October 2003, pp. 63-71.

[8] M. Turner et all., "Turning software into a service", *IEEE Computer*, October 2003, pp. 38-44.

[9] W3C- The World Wide Web Consortium, "OWL Web Ontology Language Guide" http://www.w3.org/TR/owl-guide/, access: 15 March 2005.

[10] Protégé knowledge-base editor. Stanford Medical Informatics. http://protege.stanford.edu/, access 15 March 2005.