BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume 7, No 2

Sofia • 2007

Models, Languages and Tools for Integration of Information Resources and Services*

Hristina Daskalova

Institute of Information Technologies, 1113 Sofia
E-mail: daskalovahg@abv.bg

Abstract: *The contemporary development of Web is characterized by weak structuring of the data, low level of connection between them and orientation towards human interference in the discovery and integration of Web resources in the information space. The principles, comprising Web, define the intense rates of its development and also the problems, connected with the discovery of information, altering in content and location. The integration of the data and the applications consists in the combined use of the data and processes without any serious changes in the applications or in the data structures. This is achieved applying simultaneously different technologies. After making a brief review and analysis of the existing possibilities, the purpose of the present paper is to outline the common problems, which are nowadays discussed and require future solutions, related to the integration of the information resources and services.*

Keywords: *Web-services, integration, semantic Web, languages, tools, Web technologies.*

1. Capabilities of semantic Web in Web technologies

Like any other area, the integration of information resources and services implies many technical problems, some of them could be solved or ignored, while others still remain unsolved. The technologies for combined data modeling are connected with semantic Internet, whereas the weak connected, service-oriented, communication protocols provide the infra structure for the integration of large volumes of information, available in Internet. The Internet infra structure must give meaning to the information and make it appropriate for automation, integration, its realizing and multiple use.

* The research work presented is done in conformance with the theme "Methods and tools for integration of the information resources in distributed systems and networks", No 010085.

The purpose of the paper presented is to make a short survey of the existing possibilities and also to outline some common problems, which are recently developed and which require solutions.

In order to raise the level of information integration and to provide active data processing, one alternative is the implementation of Semantic Web in Web technologies [1]. Its purpose is to raise the level of information integration, to develop data machine processing, to allow more adequate responds to clients' requests, etc. Semantic Web is based on the data model Resource Description Framework (RDF) [2] that enables combining of information from different resources. The other component of Semantic Web is RDF/XML – syntax, which represents RDF data in a XML form. The next level in the pyramid of Semantic Web technology is RDF Schema (RDFS) language, which describes the vocabulary of terms, representing the classes and the properties of Web resources in RDF.

Thus, unlike XML-schemes that describe documents structure, RDF-schemes permit the defining of the data semantics, presented with the help of XML. In this way RDFS becomes the ground enabling the use of some more developed languages in the description of models of object domains – languages, describing ontologies. The latter makes possible the application of mathematical logic and semantic data processing in Web systems. RDF data model is characterized by the fact, that the resources and their properties are identified with the help of global identifiers – URI. Thus RDF defines the object domain in terms of resources, resources properties and properties values. This data model could be regarded as a set of confirmations – a subject, a predicate (property), an object and confirmations which are presented as a graph, formed by these confirmations. RDF/XML syntax enables the storing of a similar graph in a successive record, appropriate for data exchange. This syntax also permits different forms and abbreviations in the record on one and the same graph. RDFS gives the mechanism of determining the necessary set of resources types and properties and enters some terms as classes, subclasses, properties and subproperties, as well as the possibility to imply constraints on them. This enables the defining of the resources and properties classes as elements in a vocabulary with data types and to specify what properties correspond to given classes.

RDFS expresses these vocabularies with the help of RDF tools. In this way a set of apriori given resources and properties is obtained, with the respective meaning colouring that could be used to describe new RDF vocabularies. RDFS enables defining of the unique resources, identified by URI classes, which represent the conceptual model of a certain object domain and also these, identified by URI properties. RDFS language provides the basic possibilities to describe data schemes in a certain object domain. This language has an internal mechanism of expanding, since RDF data model is used to describe schemes that could expand the description of each resource by additional information. RDF Schema is the basis for the application of the more complicated language, describing the ontologies of the object domains (Ontology Web Language (OWL) [3], which ensures the definition of more complex constraints on the classes of the properties and on the metadata structure.

As an illustration of the particular importance of the problem of information processes and systems integration, many European and world-wide projects could be pointed out, one of them being the project, in which the Institute of Information Technologies, BAS, participates – INFRAWEBS – Intelligent Framework for Generating Open (Adaptable) Development Platforms for Web-Service Enabled

Applications Using Semantic Web Technologies, Distributed Decision Support Units and Multi-Agent-Systems [4], that was completed with the development of an open software platform for Web sevices. The project development was oriented towards the use of WSMO language , which represents the European initiative for semantic Web services [5]. WSMO ensures formal ontologies and a description language for the different aspects, connected with semantic Web.

The ideology of the software platform developed – INFRAWEBS Integrate Framework contains separate semantic Web devices (Fig. 1), which provide the tools for analysis, design and support of WSMO-based Web services and give the possibility to create a new set of semantic applications, which have not been previously developed in the platforms.

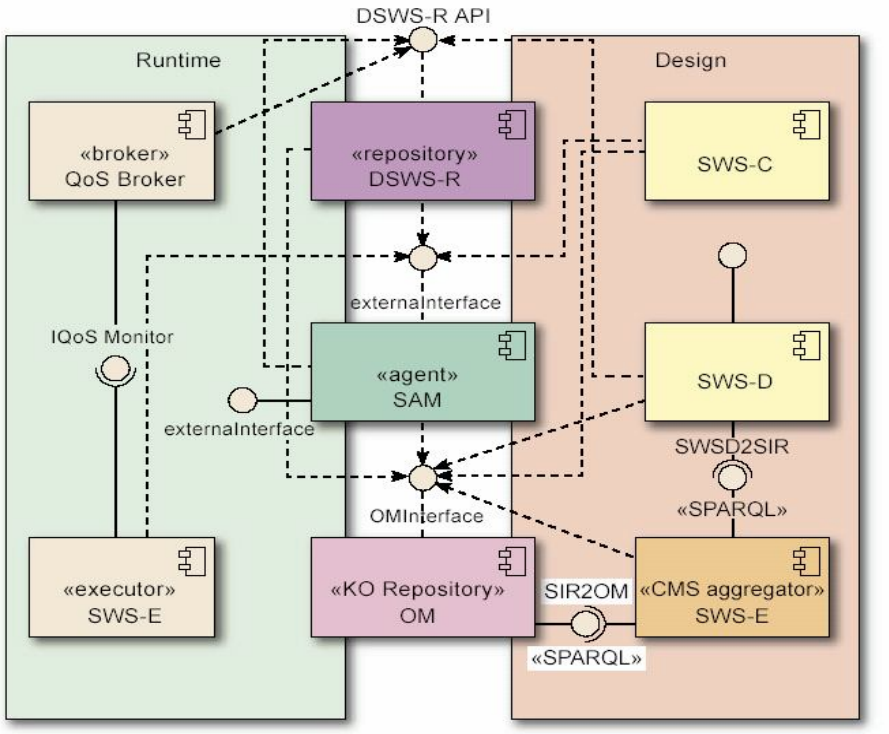

Fig. 1

The efforts, connected with the formulation and solution of the problems, concerning the integration of the information resources and services during project development, are described in some of the references presented [6-19].

2. Decentralization in information processing and storing

Decentralization is an important property of Web technologies, both in information processing and its storing. That is why it is not possible to use abbreviated names instead of URI, but the resource description could be replaced by a reference to URI.

Still, there is no certainty in the resource complete description, because it is possible to find at least one more description, giving supplementary information about this resource.

The division of the metadata schemes in successively expanding sub-schemes is possible due to data decentralization in RDF: each scheme is regarded as a set of confirmations, and the expanded scheme – as a set of additional confirmations. OWL enables the indication of the metadata for the schemes and in some cases – their functional relations, schemes importing. In this case all confirmations in the scheme imported become confirmations in the importing ontology – sub-scheme. The sub-scheme does not only define its own classes and properties, but also points to additional information about the imported classes and properties, adds new properties to the classes imported, precises the type of values and constraints.

Several different methods are used to import RDF Schema and OWL schemes: use of the mechanism of subclasses in order to indicate classes' categorization, to specify the semantics of terms and properties set; application of the mechanism of equivalence of classes, properties and values of the vocabularies elements. These mechanisms could be used to indicate the representation of the schemes on widely applied profiles of the metadata. The systems using metadata are the ground of the search in distributed data bases with the help of the network protocol Z39.50 and of server and clients' programs, designed on its basis. This technology uses GILS system in order to report the government resources of USA [20]. This technology requires the use of special clients' programs and a lock in the interaction with Web servers.

Nowadays the system called Dublin Core (DC), which uses metadata, with the complete name Dublin Core Metadata for Simple Resource Discovery, is mostly widespread. DC standard is an efficient set of elements describing a wide spectrum of network resources. Fifteen elements are included with semantics, defined by inter-disciplinary specialists in the area of librarian science, computing, texts encoding, museums and other related aspects. There exist two classes of terms in DC – elements and qualifiers which can be organized in simple confirmations. In this language the resources are regarded as objects. Each element has a limited set of qualifiers - these are attributes that might be used for some precisions, but not as expansions, of the element values. The Dublin Core Metadata Initiative (DCMI) has defined the standard ways for elements "qualification" using different types of qualifiers.

DC format aims at achieving the following characteristics:

– simplicity in resources design and support, use of a small and simple set of elements, allowing non-specialists to describe the information resources, which simplifies their discovery in the network environment;

– easy comprehension of the semantics which allows the generation of the metadata directly by the resource designers;

– international scale, the set of elements being developed in English, but at the same time there are versions in many other languages and a special group coordinates the attempts to combine these versions in a distributed register with the help of RDF technology. A lot of governments have accepted DC as a national standard for metadata.

DC format is particularly appropriate in describing objects of the type of "publications", which is determined by the following advantages of the standard:

– the set of basic semantic elements is compact, enabling setting of the necessary attributes;

– simplicity of its implementation;

– presence of a complete set of normative documents and technical support;

– the semantics of every element is specified with the help of qualifiers, both standard, known and comprehensive, and also specially developed for a given specification of the semantic meaning of a certain attribute during data exchange within a community;

– the standard possesses the possibility to use different semantic schemes, vocabularies and others, defining the mechanism that enables information extraction from descriptions, which have non-standard expansions of their domains names;

– the standard is gaining worldwide importance.

The vocabulary of DCMI type is a powerful inter-disciplinary list of approved terms that could be used as values of the element Resource Type, identifying in this way the type of the information resource. The specification Dublin Core Structured Values (DCSV) of the basic structural values of DCML is intended for storing the values of the languages attributes in HTML and XML form.

3. Meta descriptions of Web-services and importance of UDDI specification

The systems, describing Web-services, come into sight in 2000 for the purpose of some commercial projects, when companies like Microsoft, Hewlett-Packard and others create UDDI specification (Universal Description Discovery and Integration) [21]. It is a universal method for description, discovery and integration of Web-services in B2B electronic commercial systems. The main component of UDDI is a business register, which is a data base of common use, where the companies can register their service providers and present information concerning their electronic business. UDDI business register does not limit the type of services and might contain information about automatic or manual services, services in the area of transport, tourism, education, etc. But this register is designed mainly for a network and Web-services.

Using the functional options, included in UDDI, information can be discovered about the providers' services and also define the compatibility of the technologies used by the partners. The development of respective interfaces is done, if needed. Being an international standard, UDDI ensures common environment for interaction within the frames of an arbitrary infrastructure – cooperative or global. It comprises the names of the business enterprises, their mail codes, the persons for contact, telephone numbers, email addresses, URL of the Web-services suggested, metadata describing the program interfaces to Web-services, etc.

UDDI specification contains a set of rules for meta description of the Web-services with the help of XML language and also allows the discovery of the services on the basis of meta descriptions, though no special protocols or languages for data exchange are defined. The service description may be in any natural language, and for its access either the telephone number of the services provider or any of the program

protocols may be used, for example ebXML Message Service, COM+, DCOM, e-Speak, XML-RPC, CORBA, Java RMI, etc.(Fig. 2). But UDDI is mainly used with SOAP protocol and the language for services description – Web Services Description Language (WSDL) [23]. The XML descriptions, stored in UDDI register, give information about the interfaces supported by the services provider, for example whether they support API-interfaces like Sun ONE or Microsoft Visual Studio .NET, or WebLogic Workshop of BEA Systems company, or WebSphere Business of IBM, or HP Web Services Platform, taking into account the specifications SOAP/WSDL/ UDDI. After finding what API-interface is used by the provider, a connection can be realized with the Web-service selected and data are exchanged with the help of SOAP protocol [22]. The unauthorized access and probable alterations in any experimental or calculations data, threaten seriously the security. The last version of the standard includes some security options while using Web services, requiring an electronic signature, some encoding technologies, a catalog of the access permissions, etc.

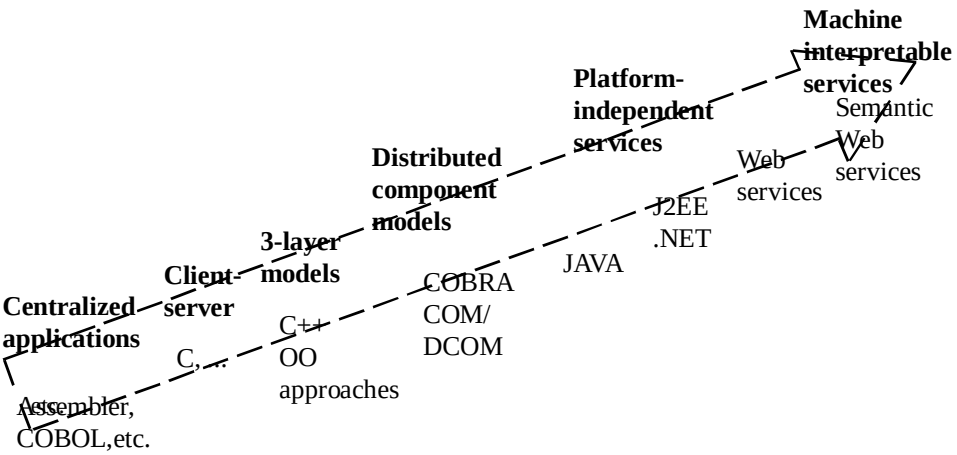

Fig. 2. Genesis of Web services towards semantic services

Microsoft, IBM, SAP and NTT companies have already designed some common UDDI nodes, combined in an united register named Universal Business Registy (UBR). UBR nodes are located on the East and West coast of North America, in Europe and Asia, and the UDDI business registers are a distributed data base. The nodes operators support a common set of program interfaces (API) and possess and serve the complete dynamic copy of the whole business register. The users may access the business registers through ordinary Internet browsers. In the general case the search in UDDI register is done by standard industrial classifications (UNSPSC, NAICS, SIC), which define exactly the type of the services offered. The library with models of services descriptions (tModel) and WSDL language play a significant role in the automatic operation with Web-services.

UDDI business register enables the search for services providers and the services themselves, but also tModel registered, which could simplify the standard description of Web- services.

4. Construction of business processes on the basis of Web-services

In parallel with the projects connected with UDDI business register there are a lot of developments on the construction of business processes on the basis of Web services [24, 25, 26]. The developers, who wish to use Web services by different providers for one and the same business process must account, besides the interfaces in the type of WSDL, also the characteristics in the interaction of these services (their choreography): continuity in Web services functioning, whether data transfer is synchronous or asynchronous, the processing of a situation when an error appears during the execution of any service and many others. For this reason Sun, BEA Systems, Intalio and other companies suggested a project for specification, that standardizes the combined operation of Web services, called Web Service Choreography Interface (WSCI), and later IBM, Microsoft and BEA Systems published the alternative document BPEL4WS (Business Process Execution Language for Web Services).

With the purpose to organize universal, independent on the functioning platform and tools, interaction of the Web services with the help of standard Internet protocols (mainly HTTP), the following key problems are still in the process of their solution. They determine the integration of the information resources and services:

– the use of the functionality of one Web service in others by its presenting with the help of the corresponding program interface – Application Programming Interface (API);

– messages formatting with the help of XML language;

– the exchange of XML messages between services with the help of the protocol for objects access - Simple Object Access Protocol (SOAP);

– publishing of Web services interfaces, using WSDL language (Web Services Description Language);

– forming the “choreography” of Web processes execution with the help of BPEL4WS, WSCI or other specifications;

– solving some legal problems connected with URI application;

– development of projects concerning the updating of the global business register.

5. Conclusions on the basis of analysis of the status quo

The information systems may have enormous volume of not-structured, not-homogenous data, with distributed nature and using different methods for control and access to the information resources and services. As a result of the survey above presented some general problems could be outlined, which are still being developed and need new solutions:

– Providing of the interaction between various information resources, which would support the respective interfaces, protocols and mechanisms for access to the information resources;

– Assuring an unified approach towards the data from information resources, that are not syntactically homogenous, use various models and forms of representation, reducing them to widespread models of data and formats;

– Introducing a sequence in information search, applying the structural presentation and metadata, that describe the resource content by the corresponding values and connections with the other resources. This enables the automatic analysis of the resource content and the efficiency in discovering information of various type;

– Design and co-ordination of standard applied profiles of the metadata and ontologies, which simplify the integration of different systems, ensuring the automatic exchange, processing and conversion of the metadata. The metadata may be related to different object domains and have different type and interpretation within one object domain. The approach of increasing the degree of metadata structuring enables their fitting to certain applications;

– Support of resources global identification, using unique identifiers, which enable the connection between the resources of different repositories in the distributed environment, combining the connected data from the repositories in virtually united resources;

– In order to decrease the network loading, the requests for any information resources must be executed by the corresponding subset of repositories, providing requests routing and answers uniting. When the information is frequently used, a replication of the metadata from smaller servers on a bigger one is suggested, concentrating the information, most frequently required, on a given number of powerful servers;

– Solving the problem for access control and avoiding the necessity for user's signing in each information resource, supporting just one entry.

6. Conclusion

The most ambitious purpose of integration would be the achieving of a common virtual system. This could assure real access to the information requested, no matter where located. In this way the data bases, the applications and other information resources would be accessible always and everywhere. Most probably in the near future some practical steps will be made and joined efforts of the researchers applied in order to accept rules and open standards; to prepare typical solutions for adapting to the applied systems and infra structures developed, which will support different levels of interacting distributed data and applications. It is also expected that information systems of general purpose will be developed, containing modular organization and larger functional possibilities.

References

1. Berners-Lee, T., James Hendler, Ora Lassila. The Semantic Web, Scientific American, May 2001.
http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21
2. RDF Vocabulary Description Language 1.0: RDF Schema, W3C Working Draft 23 January 2003.
http://www.w3.org/TR/rdf-schema
3. The OWL Web Ontology Language.
http://www.w3.org/TR/2004/REC-owl-features-20040210

4. EU Project INFRAWEBS – IST FP62003/IST/2.3.2.3 Research Project No 511723.
5. **http://www.wsmo.org/TR/d2/v1.1/20050210/**
6. D a s k a l o v a, H. Web Integration For Enterprise Applications. – In: Proc. of the Conference Euromedia-2005, 11-12 April, 2005, Toulouse, France, 207-209.
7. D a s k a l o v a, H., V. G r i g o r o v a, T. A t a n a s s o v a. Existing Approaches for Modeling and Integration of Business Logics and Processes. – In: Proc. of the 11th Annual Scientific Conference on Web Technology, New Media, Communications and Telematics Theory, Methods, Tools and Applications – Euromedia'2005, 11-13 April 2005, Toulouse, France, 210-216.
8. D a s k a l o v a, H., V. G r i g o r o v a, T. A t a n a s s o v a. Possibilities for Modeling and Integration of Business Processes. – Cybernetics and Information Technologies, Vol. **5**, No 1, Sofia, 2005, 35-44.
9. D a s k a l o v a, H., T. A t a n a s s o v a, N. B a k a n o v a. Integration of Information Resources from Distributed Repositories by Business Processes Models. – Distributed Computer and Communication Networks. – In: International Workshop, Proceedings, Sofia, Bulgaria, April 23-29, 2005, Technosphera, Moscow, 36-43.
10. A t a n a s s o v a, T., G. A g r e, H. J. N e r n. INFRAWEBS Semantic Web Unit For Design and Composition Of Semantic Web Services. – In: Proc. of the Conference Euromedia-2005, 11-12 April, 2005, Toulouse, France, 216-220.
11. N e r n, J., T. A t a n a s o v a, G. A g r e, A. M i c s i k, L. K o v a c s, J. S a a r e l a, T. W e s t-k a e m p e r. Semantic Web Service Development on the Base of Knowledge Management Layer – INFRAWEBS Approach. – In: Proc. Third International Conference Tech-2005, Varna, Bulgaria, 27-30 June 2005, FOI-COMMERCE, ISBN 954-16-0034-4, 217-223.
12. D a s k a l o v a, H., T. A t a n a s s o v a. Semantic Web Services – Where We Are and Where We Are Going. – In: Proc. of the Int. Conference "Generic issues for knowledge technology", Budapest, 14.09.2005.
13. A g r e, G., T. A t a n a s s o v a, H. J o a c h i m N e r n. Migrating From Web Services to Semantic Web Services: Infrawebs Approach. – In: Proc. of the Conference Euromedia-2005, 11-12 April, 2005, Toulouse, France, 221-225.
14. A t a n a s s o v a, T., H. D a s k a l o v a, N. B a k a n o v a. Design Aspects of Application Integration in Distributed Information Systems. – In: Proc. of International Workshop, Distributed Computer and Communication Networks. Sofia, Bulgaria, October 30-November 2, 2006, 27-37, Moskow.
15. D a s k a l o v a, H., T. A t a n a s s o v a. On the Requirements to the Methods for Web Service Composition. – In: Proc. of 12th Annual Scientific Conference on Web Technology, New Media, Communications and Telematics Theory, Methods, Tools and Applications EUROMEDIA 2006, April 17-19, 2006, Athens, Greece,115-121.
16. A t a n a s s o v a, T., H. D a s k a l o v a. An Approach For Semantic Web Service Composition. – In: Proc. of 12th Annual Scientific Conference on Web Technology, New Media, Communications and Telematics Theory, Methods, Tools and Applications EUROMEDIA 2006, April 17-19, 2006, Athens, Greece,122-129.
17. D a s k a l o v a, H., V. G r i g o r o v a. A Survey on the Integration of Enterprise Applications as a Service. – International Journal "Information Theories and Applications", 179-183, 2006.
18. D a s k a l o v a, H., T. A t a n a s o v a. Web Services and Tools for their Composition. Considering Aspects of Digital TV Workflow. – In: Proc. of 13th Annual Scientific Conference on Web Technology, New Media, Communications and Telematics Theory, Methods, Tools and Applications EUROMEDIA 2007, April 25-27, 2007, Delft, The Netherlands,139-144.
19. D a s k a l o v a, H., T. A t a n a s s o v a, N. B a k a n o v a. Hybrid Applications: Using and Combining of Web Services. – In: International Workshop, Distributed Computer and Communication Networks Proceedings, Vol. **2**, Moskow, Russia, September 10-12, 2007, 91-97.
20. Application profile for the Government Information Locator Service (GILS). **www.gils.net/prof_v2**

21. UDDI Version 3.0.
http://uddi.org/pubs/uddi-v3.00-published-20020719.htm
22. SOAP Version 1.2 Part 1: Messaging Framework.
http://www.w3.org/TR/2003/REC-soap12-part1-20030624
23. Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language.
http://www.w3.org/TR/2004/WD-wsdl20-20040326/
24. Web Services for Business Process Design.
http://www.gotdotnet.com/team/xml_wsspecs/xlang-c/,
http://xml.coverpages.org/xlang.html
25. Business Process Execution Language for Web Services Version 1.1.
http://www-106.ibm.com/developerworks/library/ws-bpel/
26. Web Services Flow Language.
http://www-3.ibm.com/software/solutions/Webservices/pdf/WSFL.pdf,
http://xml.coverpages.org/wsfl.html, http://www.ebpml.org/wsfl.htm