SEMANTIC WEB-BASED DISCOVERING OF HYPERMEDIA RESOURCES

Sabin-Corneliu Buraga

Faculty of Computer Science, "Al.I.Cuza" University of Iaşi, Romania
E-Mail: *busaco@infoiasi.ro*, WWW: *http://www.infoiasi.ro/~busaco/*

Abstract: The paper presents a semantic Web-based architecture used to discover hypermedia resources. This platform is based on Web services and agents, exploiting the spatial/temporal relations related to Web resources. All involved information within the system is XML-based, by using special languages for expressing metadata and temporal relationships between Web hypermedia documents.

1. INTRODUCTION

Computers are used principally to render this hypermedia information, not to reason about it. Information retrieval has become ubiquitous with the WWW's development and information needs no longer to be intended for human readers only, but also for machine processing, enabling intelligent information services, personalized Web sites, and semantically empowered search engines – this is the seminal idea of the *Semantic Web* [20, 28].

This paper will present *ITW* – a semantic Web-based distributed platform for hypermedia resource discovery. The general architecture of the *ITW* system uses software agents, Web services, and other software entities such as *CGI (Common Gateway Interface)* [17] scripts. The platform takes advantage of the temporal relations established between Web sites' resources and uses a semantic Web-based model for the representation of metadata and additional information that involves time, by using XML and RDF syntactic constructs.

2. GENERAL ARCHITECTURE OF THE *ITW* SYSTEM

2.1 General Presentation

The main purpose of *ITW* system is to offer a heterogeneous interoperable infrastructure [15], based on Web components, in order to discover hypermedia resources. Using a Web interface, the user will be able to make complex queries that involve time. The information and the associated *RDF (Resource Description Framework)* [22] metadata generated by the *ITW* system will be stored on independent Web servers. Even if one server is shutdown, the system will be able to continue its execution, providing the same capabilities.

In fact, *ITW* can be considered as a multi-language and multi-platform architecture for (hypermedia) resource discovery based on semantic information related to the Web resources.

In the resource discovery process, one of the important issues is to deal with time. Using our defined RDF-based model, the *ITW* software agents will be able to reason about the spatial and/or temporal relations established between certain resources localized in different Web sites.

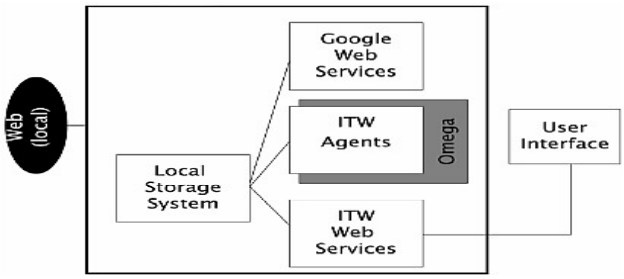

Fig.1 General Architecture of the *ITW* System

The general architecture of the *ITW* system consists of two main entities (see Fig. 1):

- *ITW Web services* – their role is to give preferred information about the resources and the access to these resources; the Web services can be invoked by diverse software applications (e.g., user agents – such as Web browsers or special clients –, other agents or Web services). For these Web services, we could associate semantic descriptions specified in *OWL (Web Ontology Language)* [28];
- *ITW agents* – their role is to effectively discover distributed multimedia resources stored on different sites; these agents are intended to be implemented within a multi-agent system, such as Omega [11] or ADF (Agent Development Framework) [25].

2.2 ITW Services

The Web services part of the application consists of some local Web services to give information about the resources stored on the local Web (i.e. the intranet or the public Web site), external Web services (offered, for example, by Google) and a local storage system for storing metadata and relations between found resources. We can think these Web services as one entity (similar with a Grid [21]).

These services can be independently invoked by other external entities. The information exchanged by Web services and their clients is stored as *Simple Object Access Protocol (SOAP)* messages [28, 29]. For each local design Web services, the system provides *Web Services Description Language (WSDL)* [29] descriptions and uses the available WSDL documents in order to invoke external Web services. To add semantic descriptions for each *ITW* service, we suggest the use of DAML-S/OWL-S specification (see [10]).

The *ITW* services employ RDF documents about the discovered hypermedia resources. These documents are automatically produced by the *ITW* agents (presented in section 2.3) and are stored within the storage system.

2.3 ITW Web Agents

The ITW Web agents are intended to be developed within a multi-agent platform, called *Omega* [3, 11]. To enable the flexible querying and accessing mechanisms about the distributed Web resources, *Omega* offers a facility for serialization – in an independent manner – the data and metadata (objects) processed by the system. The serialization method is detailed in [3] and uses XML Schema and SOAP. Additionally, for each object, different metadata constructs are attached to specify several semantic properties [11, 14]. These descriptions are using RDF model.

The agents of the *Omega* system have the following tasks to be accomplished in their activity of discovering multimedia resources on Web:

- Using different XML constructs expressed in our defined *XFiles* [7] and *TRSL (Temporal Relation Specification Language*) [14] languages, for each Web resource an RDF document is produced (details about this process are given in [9]). The *XFiles* documents are used to keep all significant metadata that can be associated with a Web resource: location, type (e.g., XHTML document, PNG image, JavaScript program etc.), access way, timestamp of last modification and others (see [7]). The TRSL constructs are used to store temporal information regarding the relationship between resources, by using the *ITL (Interval Temporal Logic)* [4-6] formal model.

- Using RDF constructs regarding the temporal relations, one key aim is to preserve these relations (e.g., if a Web resource is in an ITL relation *Before* with another one, the agent will try to maintain this relation, by checking regularly the metadata associated with the involved resources). For this relation, the user could specify certain actions to be executed by using TRSL constructs.

- The internal behavior of the multi-agent environment can be modeled by BDI^{K}_{CTL} logic [24, 26], often used in the context of multi-agent systems [10].

For a suitable communication between agents, the *ITW* system uses our XML-based agent-communication language over SOAP messages, detailed in [11].

2.4 ITW User-Interface

The user-interface consists of one CGI script that uses *Extensible Markup Language (XUL)* [23] in order to provide a flexible query user-interface. To describe user interactions on different controls of the Web interface, a protocol must be adopted. One of the best solutions is to use the *XUP (Extensible User-interface Protocol)* [28].

The user queries are stored into a customized version of our defined *WQFL (Web Query Formulating Language)* [16] language, to indicate supplementary information about the search (e.g., relation with another resource, method of access, resource type, etc.).

In [12] and [13], we investigated the use of Perl-like regular expressions to give a more flexible technique for query formulating.

3. AN IMPLEMENTATION PROTOTYPE

According to general designed architecture described in section 2, the actual implementation prototype of the *ITW* system includes the subsequent components:

- One Perl script that generates the XUL user-interface and functions as a basic client for the involved hypermedia resource discovery services. Using *Extensible Stylesheet Language (XSL)* [1, 28], the XUL documents can be transformed in other XML-based languages, such as XHTML or WML, in order to support all actual Web browsers;
- Two ITW Web services available on Linux and Windows platforms; one Web service is implemented in Perl [17] by using *SOAP::Lite* module and Apache Web server; the other is implemented in C# language on Microsoft .NET Framework, using *IIS (Internet Information Services)* Web server;
- One external Web service freely provided by the Google search engine in order to discover world-wide hypermedia Web resources.

The implemented elements were tested on Windows XP – using .NET Framework 1.1 – and different Linux distributions (Red Hat, Fedora Core, Mandrake, and Mandriva) – using Perl 5.8 and Apache 2. For XML processing, the *XML::Parser* module were used. The *ITW* agents are implemented in C++ language. The storage system consists of two open-source relational database servers: MySQL (under Windows) and PostgreSQL (under Linux).

The *ITW* system can be viewed as a semantic Web-based platform for discovering hypermedia information within the organizational intranet, such as an enterprise Web portal – see [18] and [19]. Another use of the system is in the context of e-learning [8].

4. CONCLUSIONS AND FURTHER WORK

A heterogeneous architecture, called *ITW*, for hypermedia resource discovery was presented. This system consists of a collection of Web services, software agents and other components and uses a RDF/XML model (introduced in [14]) for semantic representation of metadata and spatial and temporal relations between hypermedia resources available on Web.

The presented implementation was focused on the development of the local Web services – deployed on Windows and Linux platforms. For remote access to resources, the Google's search Web service is used. To increase the processing speed, for the next version we intend to store the XML/RDF constructs within an XML native database, such as Apache XIndice [27]. Also, other

further goal is to incorporate the *ITW* system into the *tuBiG* Grid-oriented architecture [2], to put our project in the context of Grid computing [21].

REFERENCES

[1] S. Adler *et al.*, *Extensible Stylesheet Language (XSL) Version 1.0*, W3C Recommendation, Boston, 2001: *http://www.w3.org/TR/xsl/*

[2] L. Alboaie, S. C. Buraga, S. Alboaie, "tuBiG – A Layered Infrastructure to Provide Support for Grid Functionalities", in M. Paprzycki (ed.), *Proceedings of the 2nd International Symposium on Parallel and Distributed Computing*, IEEE Press, 2003

[3] S. Alboaie, S. C. Buraga, L. Alboaie, An XML-based Serialization of Information Exchanged by Software Agents", International Informatica Journal}, vol.28, no.1, April 2004

[4] J. Allen, "Time and Time Again: The Many Ways to Represent Time", *International Journal of Intelligent Systems*, 6 (4), 1991

[5] J. Allen, "Maintaining Knowledge about Temporal Intervals", *Communications of the ACM*, 26 (11), 1983

[6] J. Allen, P. Hayes, "Moments and Points in an Interval-based Temporal Logic", *Computational Intelligence*, 5 (4), 1989

[7] S. C. Buraga, "A Model for Accessing Resources of the Distributed File Systems", in D. Grigoraş *et al.* (eds.), *Advanced Environments, Tools, and Applications for Cluster Computing. NATO Advanced Research Workshop,* Lecture Notes in Computer Science – LNCS 2326, Springer-Verlag, Berlin, 2002

[8] S. C. Buraga, "Development Agent-Oriented E-learning Systems", in I. Dumitrache, C. Buiu (eds.), *Proceedings of the 14th International Conference on Control Systems and Computer Science – CSCS14*, Politehnica Press, Bucureşti, 2003

[9] S. C. Buraga, "ITW – An Architecture based on Distributed Web Components for Multimedia Resource Discovery", *Scientific Annals of the "Al. I. Cuza" University of Iaşi*, Computer Science Series, tome XIII, 2003

[10] S. C. Buraga, *Semantic Web* (in Romanian), Matrix Rom, Bucureşti, 2004

[11] S. C. Buraga, S. Alboaie, L. Alboaie, "An XML/RDF-based Proposal to Exchange Information within a Multi-Agent System", in D. Grigoraş *et al.* (eds.), *Proceedings of NATO Advanced Research Workshop on Concurrent Information Processing and Computing*, IOS Press, 2005 (to appear)

[12] S. C. Buraga, M. Brut, "A Proposal for a Web Structural Search Language Based on XML Technologies", *Scientific Annals of the "Al. I. Cuza" University of Iaşi*, Computer Science Series, tome X, 2001

[13] S. C. Buraga, M. Brut, "Different XML-based Search Techniques on Web", *Transactions on Automatic Control and Computer Science*, vol. 47 (61), No. 2, Politehnica Press, Timişoara, 2002

[14] S. C. Buraga, G. Ciobanu, "A RDF-based Model for Expressing Spatio-Temporal Relations Between Web Sites", in *Proceedings of the 3rd International Conference on Web Information Systems Engineering – WISE*, IEEE Press, 2002

[15] S. C. Buraga, P. Găbureanu, "A Distributed Platform based on Web Services for Multimedia Resource Discovery", in M. Paprzycki (ed.), *Proceedings of the 2nd International Symposium on Parallel and Distributed Computing*, IEEE Press, 2003

[16] S. C. Buraga, T. Rusu, "An XML-based Query Language Used in Structural Search Activity on Web", *Transactions on Automatic Control and Computer Science*, vol. 45 (59), No. 3, Politehnica Press, Timişoara, 2000

[17] S. C. Buraga *et al*., *Web Programming in bash and Perl* (in Romanian), Polirom, Iaşi, 2002

[18] M. Cioca, S. C. Buraga, "New Tools for Human Resource Management in e-Business: Combining UML Language, Reference Architectures and Web Programming", *IEEE International Conference on Industrial Informatics – INDIN'03 Proceedings*, IEEE Press, 2003

[19] M. Cioca, S. C. Buraga, "Instruments and Web Technologies for Implementing Architectures and Integration Informatics Systems in Virtual Enterprises'', *CD-ROM Proceedings of the 3rd International Conference on Research and Development in Mechanical Industry – RaDMI 2003*, Herceg Novi, Montenegro Adriatic, 2003

[20] J. Davies, D. Fensel, F. van Harmelen (eds.), *Towards the Semantic Web*, John Wiley & Sons, England, 2003

[21] I. Foster, C. Kesselman (eds.), *The Grid: Blueprint for a Future Computing Infrastructure*, Morgan Kauffmann Publishers, 1999

[22] F. Manola, E. Miller (eds.), *RDF Primer*, W3C Recommendation, Boston, 2004: *http://www.w3.org/TR/rdf-primer/*

[23] I. Oeschger, *XUL Programmer's Reference Manual*, Mozilla.Org: *http://www.mozilla.org/xpfe/Xulref.zip*

[24] A. Rao *et al.*, *Formal Methods and Decision Procedures for Multi-Agent Systems*, Technical Report No. 61, Australian Artificial Intelligence Institute, 1995

[25] O. Nichifor, S. Buraga, "ADF - Abstract Framework for Developing Mobile Agents", in D.Petcu *et al.* (eds.), *Proceedings of the 6th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing - SYNASC 2004*, Mirton Publishing House, Timişoara, 2004

[26] M. Wooldridge, N. Jennings, "Intelligent Agents: Theory and Practice", *Knowledge Engineering Review*, 1995

[27] * * *, *XML Apache Project*, 2005: *http://xml.apache.org/*

[28] * * *, *World Wide Consortium's Technical Reports*, Boston, 2005: *http://www.w3.org/TR/*

[29] * * *, *Web Services*: *http://www.webservices.org/*