Next Generation Web Technologies in Content Management

Norberto Fernández-García
Telematic Engineering Department
Carlos III University of Madrid
Avda. Universidad, 30
Leganés, Madrid, Spain
berto@it.uc3m.es

Luis Sánchez-Fernández
Telematic Engineering Department
Carlos III University of Madrid
Avda. Universidad, 30
Leganés, Madrid, Spain
luis@it.uc3m.es

Jesús Villamor-Lugo
Telematic Engineering Department
Carlos III University of Madrid
Avda. Universidad, 30
Leganés, Madrid, Spain
jvl@it.uc3m.es

ABSTRACT

The development of information and communication technologies and the expansion of the Internet means that, nowadays, there are huge amounts of information available via these emergent media. A number of content management systems have appeared which aim to support the management of these large amounts of content. Most of these systems do not support collaboration among several, distributed sources of managed content. In this paper we present a proposal for an architecture, Infoflex, for the efficient and flexible management of distributed content using Next Generation Web Technologies: Web Services and Semantic Web facilities.

Categories and Subject Descriptors

H.3.5 [**Information Systems**]: Information Storage and Retrieval-Online Information Services

General Terms

Design

Keywords

Content Management, Web Services, Semantic Web

1. INTRODUCTION

At different application domains, there exists the need to offer a service that permits managing, producing and offering quality information. The entities devoted to the production and/or the storage of contents must face several requirements. Some of them are:

- Content creation and manipulation management.
- User Profile Management.
- Decision criteria on the contents to be published.
- Integration of several fragments in one document (i.e. for constructing a web page from several content sources).
- Validation of published contents.

In this paper, we show an architecture (currently at development process) for the management of flexible and distributed contents: the **Infoflex System**. The objective of such system is to provide all the required facilities of up-to-date Content Management Systems (CMSs), while improving their capabilities for accessing distributed and heterogeneous content sources. As far as we know, the novelty of our approach is the use of Web Services and Semantic Web technologies to allow a client (human or computer system) to retrieve information from different distributed and heterogeneous CMSs and to perform a unique query on several of such systems.

This paper is organised as follows. Section 2 introduces how Web Services are used in our framework. Section 3 introduces the uses of Semantic Web facilities in Infoflex. Section 4 describes the proposed working model and architecture. Concluding remarks complete this paper.

WWW2004, May 17–22, 2004, New York, New York, USA.
ACM 1-58113-912-8/04/0005.

2. WEB SERVICES IN INFOFLEX

Web Services [7] have been defined as self-describing applications that can discover and engage other web applications to complete complex tasks over the Internet. This technology was designed to ease the cooperation among remote applications.

The interoperability model proposed by Web Services perfectly fits our distributed contents access requirements. The choice of Web Services as communication mechanism provides several advantages:

Interoperability and platform independence because Web Services provide a standard text-based communication mechanism.

Service location we can use the service location facilities of Web Services to help clients to locate content sources in a dynamic manner.

Service cooperation using BPEL4WS (Business Process Execution Language for Web Services) [2], we can define complex interactions based on the coordination and cooperation of different Web Services. This way, complex tasks can be performed easily.

3. SEMANTIC WEB IN INFOFLEX

The Semantic Web [3] was born with the idea of making information "understandable" by machines. Basically this requires to describe data using metadata and define the semantics of such metadata in ontologies.

In the scope of the Infoflex system, Semantic Web facilities are used to deal with heterogeneity of content sources (differences in

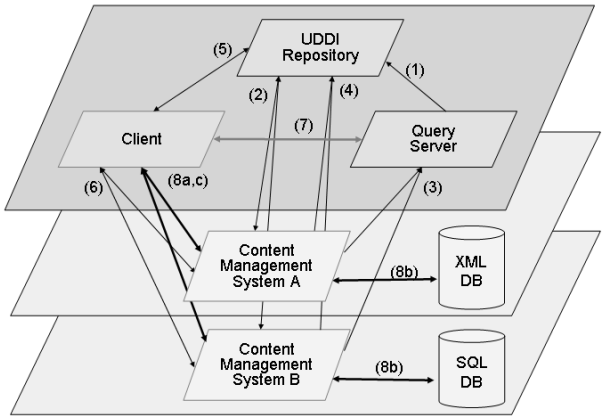

Figure 1: Architecture and Working Model

data models, in data storage formats, etc). We propose to annotate the information of each content source with metadata (using for example RDF [6]). This metadata will be described by a content provider specific ontology (written in OWL [5]). To achieve the interoperation of different sources, modelled with different ontologies, we must define a general, consensual ontology and mapping rules between this general ontology and each specific one. A special kind of Web Service, that we call **Query Server** (QS) will store the ontologies and the mapping information. This service will allow semantic translation between general and provider specific ontologies.

Other approaches to use ontologies in integration of heterogeneous data sources can be found at [4, 1].

4. INFOFLEX WORKING MODEL

In this section we will introduce the proposed working model of our Infoflex framework. The main actors involved in the full process are:

Clients who request information using a specific query language. This language has not been defined yet. It will be XML based.

Query Servers which translate queries from the model represented in the general ontology (used by the Client to make his queries) to the CMS information model.

Content Management Systems (or Content Providers) which store and manage the requested information.

UDDI repositories where all the Web Services involved on this scenario are registered.

In figure 1, we can see the steps which are necessary to obtain information from one provider. These are:

1. The Query Server registers two Web Service interfaces (one for clients, other for CMSs access) into an UDDI repository.
2. The CMS access the UDDI looking for the Query Server.
3. The CMS registers its ontology in the Query Server.
4. The CMS generates dynamically a BPEL4WS document. This document specifies all the steps (WS invocations) that Clients must follow to query the CMS. In our proposal, the document is obtained by the Clients using a WS, which is registered by the CMS in a UDDI.
5. The Client access the UDDI looking for the CMS. It obtains the reference of its WS.
6. The Client invokes the Web Service of the CMS. It obtains as the result the BPEL4WS document. The Client executes this document to retrieve the desired information.
7. During the execution, the Client writes a query, using concepts in the general ontology. The BPEL4WS execution engine uses this query (as is indicated in the BPEL4WS document) to invoke the QS, which translates it to the CMS model.
8. The obtained translated query is automatically used by the BPEL4WS engine to invoke the CMS and retrieve the desired information (if it is avaliable).

We have implemented a first prototype of Infoflex system achieving good results in interoperation of a client with an XML/XPath based content provider and a Relational/SQL based content provider.

5. CONCLUSIONS

In this paper we have presented a proposal of an architecture for the management of contents in a flexible and distributed manner. This proposal is based on advanced Web technologies such as Web Services or Semantic Web ones. We expect for this system to be useful at environments like business to business portals, news agencies or content syndication sites.

6. ACKNOWLEDGEMENTS

This work has been partially financed by the “Ministerio de Ciencia y Tecnología de España” (Science and Technology Ministry of Spain) through the TIC2003-07208 “InfoFlex” Project.

7. REFERENCES

[1] Amann, B., Beeri, C., Fundulaki, I. & Scholl, M. Ontology Based Integration of XML Web Resources. ISWC 2002, LNCS 2342, pp 117-131, 2002.

[2] Andrews, T., Curbera, F., Dholakia, H., et al. Business Process Execution Language for Web Services. Version 1.1 Specification. 2003.

[3] Berners Lee, T., Hendler, J., Lassila, O. The Semantic Web. Scientific American. Vol 284, n 5. 2001. pp 34-43.

[4] Maier, A., Schnurr H.& Sure, Y. Ontology-Based Information Integration in the Automotive Industry. ISWC 2003, LNCS 2870, pp. 897-912, 2003.

[5] World Wide Web Consortium. OWL Ontology Web Language Guide. http://www.w3.org/TR/owl-guide/

[6] World Wide Web Consortium. Resource Description Framework (RDF). http://www.w3.org/RDF/

[7] World Wide Web Consortium. Web Services Activity. http://www.w3.org/2002/ws/