Decomate: Unified Access to Globally Distributed Libraries

Thomas Place and Jeroen Hoppenbrouwers

Tilburg University, PO Box 90153, NL-5000 LE Tilburg, The Netherlands
{place,hoppie}@kub.nl

Abstract. The Decomate project enables mutual access to heterogeneous, distributed, and pooled digital resources of consortium members. Using a mediator architecture with a Broker and several back-end servers, a scalable and flexible system has been developed that is going in production in major European universities. Ongoing work focuses on access improvements using graphical browsing and thesaurus integration.

1 Introduction

There are several approaches to the integration of distributed, heterogeneous information sources. One approach is the implementation of architecture based on mediators [6]. In the EU funded Decomate II project [1], this architecture was used for the creation of the European Digital Library for Economics with mutual access to the heterogeneous distributed and pooled digital resources of the consortium members (the libraries of the Autonomous University of Barcelona, the European University Institute in Florence, the London School of Economics, and Tilburg University in the Netherlands). In July 2000, the project will be finished with mediator systems running at (at least) four sites in Europe giving unified access to library catalogues, article databases with links to the electronic full text, journal databases, pre-print archives, abstract and index databases, search engines indexing Web sites, and thesauri.

The content to which a mediator gives access depends on the licenses a site (a library) has acquired. Although acquiring licenses was one of the objectives of the project [1, 5], this paper and the demonstration will focus on the software that was developed by the project. The software will be made available as Open Source and can be used for other disciplines than Economics; e.g., the Decomate software is used for accessing distributed image and video banks.

2 The Mediator Architecture of Decomate

In Decomate, the mediator is implemented by three sub-systems: the Broker, the Multi-Protocol Server (MPS), and the Result Optimiser (RO). The Broker, the MPS, and the RO run as separate servers that communicate with each other by using an XML-based application protocol on top of the TCP/IP communication layer. The protocol is named XREP: XML Request and Response Protocol.

2.1 The Broker

The Broker is the interface to the Decomate system for the client systems used by the end users (Web browsers or Java applets). It handles user requests, sends search and retrieval requests to the MPS, and generates XML or HTML documents to be displayed at the desktop of the user.

The Broker is designed as a XML template resolver [4]. Each request type that can be sent to the Broker has a corresponding XML template. Resolution is done by substituting XML elements by the search and retrieval results. The Broker can support several 'sites' (user interfaces). A site is a set of XML templates, organised in a hierarchy of domains and sub-domains. A (sub-)domain consists of sub-domains and XML templates. The XML templates are the leaves of a tree that represents a site.

A special site is the SiteBuilder with which other sites including itself (bootstrapping) can be built and maintained. With the SiteBuilder, XML chunks can be created that can be shared by several XML templates. Chunks defined at the level of a (sub-)domain are inherited by its sub-domains and the XML templates that belong to the domain. In this way, the Broker in combination with the SiteBuilder implements a highly customisable, flexible, and easily maintainable user interface generator.

2.2 The Multi-Protocol Server (MPS)

The Multi-Protocol Server is the search and retrieval client of the distributed databases and document servers. It is controlled by the Broker that sends its XREP requests to the MPS.

Several protocols are supported by the MPS. At this moment, the MPS interoperates with Z39.50 servers and with document servers using HTTP. The MPS has an API for adding other protocols. E.g., for integrating Circulation Control functions of the OPAC of the Dutch library system vendor Pica, support of the Pica3 protocol was added to the MPS.

The MPS is a multithreaded server that allows for simultaneous (parallel) searching in databases distributed over the Internet.

Semantic interoperability is also taken care of by the MPS. The MPS is aware of the differences in query language and record syntax used by the databases that are accessed by the system. Records retrieved from the external databases are mapped to an internal element set.

2.3 The Result Optimiser (RO)

The Broker can decide to use the Result Optimiser (RO) for merging, de-duplicating, and (relevance) ranking (see also section 5) result sets. For the Broker, the RO acts exactly like an MPS with added functionality. Using several 'fuzzy hash' functions [3], the RO groups records together that have a high degree of overlap, but do not need to be exactly alike. The architecture of Decomate also makes it possible to integrate advanced IR techniques with standard library databases [2].

3 Personalised Services

One of the design goals of the Decomate system was to make personalised services possible. In the present system, a Current Awareness Service and a document delivery facility is implemented.

3.1 Current Awareness

The user can ask the Broker to store queries as interest profiles. For this the Broker interacts with the Current Awareness Server (CAS), using the XREP protocol (see above). The CAS uses a relational database management system to maintain the interest profiles of the users.

The CAS Robot runs the interest profiles on a daily, weekly, or monthly basis. The frequency is determined by the user. Running an interest profile means that the CAS Robot queries the new additions to the databases. The CAS Robot uses the Broker and the MPS as mediator. The results of a run are stored in the CAS repository that is implemented as a Z39.50 database. The user is notified by e-mail of the new publications that satisfy his or her interest profile, and has access to the new results via the Broker. The results are presented to the user as issues of his or her 'personal journals.' By accessing the new issues of the personal journals via the Broker, the full functionality of the system with respect to access to the electronic full text and ordering documents is available to the user.

3.2 Document Requester

When the full text of a publication is not available on one of the document servers, the user must be given the option to order it via a document delivery service. For this the Decomate system includes a server that is called the Document Requester. The Broker can send a request to the Document Requester that is interfaced to one or more document delivery services. Which services are used depends on the library that is running the Decomate system. The Document Requester has an API that allows a library to easily add interfaces to its local document delivery services.

4 Authentication and Authorisation

For personalised services the system must know the identity of a user and whether he or she is authorised for the service. The same applies for giving access to copyrighted material (including pay-per-view access for which the user is sent a bill).

The Broker mediates between the user and the local systems for authentication and directory services (information about the users: names, addresses, email addresses, affiliations) via the Authentication broker. Different authentication mechanisms can be used in Decomate: e.g., NT authentication using Samba and userid/password validation by a LDAP server or a dedicated password daemon. Preferably, a mechanism that allows for single-logon is used: the user logs on the network once and after that all services on the network are available to him or her. LDAP is the preferred protocol for directory services, but other implementations are supported.

5 The Concept Browser

In itself, searching via the Decomate system is rather standard. However, the project has defined a work package for the development of advanced access to federated digital libraries [3]. A special graphical tool has been developed (the Concept Browser) which allows navigation through the conceptual space of a discipline. Such a conceptual network consists of several loosely coupled vocabularies and thesauri that are implemented as Z39.50 databases using the Zthes profile for thesaurus navigation [7].

For the digital library for Economics, several vocabularies and thesauri are available, such as the commonly used JEL classification codes of the Journal of Economic Literature, and thesauri maintained by the project partners. The links between related concepts are graphically displayed and the links can be followed from concept to concept. Relevant concepts can be marked and irrelevant concepts can be crossed out. While navigating through the conceptual space, the information need of the user becomes known to the system. After the user is finished with navigating the conceptual space, the system generates an optimised query that is passed to the relevant databases. The information about relevant and irrelevant concepts is used by the Result Optimiser (see section 2.3) to rank the results.

The Concept Browser is implemented as a Java applet that interacts with the rest of the system via the Broker.

References

1. Decomate II project: `http://www.bib.uab.es/decomate2`
2. Hoppenbrouwers, J. and Paijmans, H.: Invading the Fortress: How to Besiege Reinforced Information Bunkers. In: *Proceedings of the IEEE Advances in Digital Libraries 2000 (ADL2000)*. IEEE Computer Society. 27–35.
3. Hoppenbrouwers, J.: Optimising Result Sets. In: *Proceedings of the Decomate Final Conference, June 22–23, 2000, Barcelona*.
   `http://www.bib.uab.es/decomate2`
4. Kristensen, A.: Template Resolution in XML/HTML. *Computer Networks and ISDN Systems* **30** (1998) 239–249.
5. Place, T.: Developing a European Digital Library for Economics: the Decomate II Project. *Serials* **12** (2) (1999) 119–124.
6. Wiederhold, G.: Mediators in the Architecture of Future Information Systems. *IEEE Computer* **25** (3) (1992) 38–49.
7. Zthes: a Z39.50 Profile for Thesaurus Navigation.
   `http://lcweb.loc.giv/z3950/agency/profiles/`