The Digital Library Information Mediator Layer

Conference Paper · January 2006

4 authors, including:

The Digital Library Information Mediator Layer

Leonardo Candela, Donatella Castelli, Pasquale Pagano, and Manuele Simi

Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" - CNR
Via G. Moruzzi, 1 - 56124 PISA - Italy
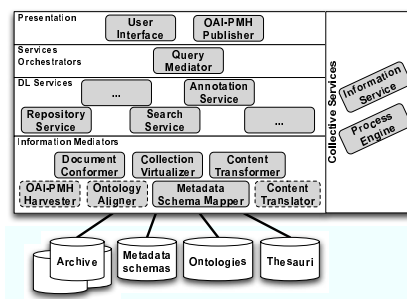{candela, castelli, pagano, simi}@isti.cnr.it

1 Introduction

Digital libraries (DLs) are often built by exploiting already existing resources. This approach reduces both the cost of the DL and the time spent for its development. The most frequently shared resources are the archives of documents, but many other type of resources, like authority files[1], thesauri[2], language dependent resources, ontologies[3], classification systems[4], gazetteers[5], can actually be re-used when building a DL.

These resources, that are created and maintained by third-parties, are usually heterogeneous, and their structure and format rarely satisfy the rules imposed by the protocol of the services that implement the end-users functionality.

Further, even if using third-party information resources for building the information space of DLs lowers the setting up costs, the resulting information space turns out to be composed by large volumes of heterogeneous data, thus rendering the DL services either too generic or not suitable to satisfy the needs of different user communities.

In order to deal with this heterogeneity the DL system must implement appropriate mechanisms. Currently, there is no standardized solution on how to design these mechanisms. Each system adopts its own approach thus duplicating efforts and excluding the possibility of re-using existing software.

We have dealt with these issues within a larger effort dedicated to define an architectural reference model for DLs[6]. We have identified the need of introducing in every DL architecture an *Information Mediator Layer*. The purpose of this functional layer is to make the

Fig. 1. The DL Layered Architecture.

---

[1] http://authorities.loc.gov/
[2] http://education.yahoo.com/reference/thesaurus/
[3] http://www.daml.org/ontologies/
[4] http://www.oclc.org/dewey/
[5] http://middleware.alexandria.ucsb.edu/client/gaz/adl/index.jsp
[6] The work described here is part of a larger effort dedicated to introduce an architectural reference model for DLs.

information disseminated by the original third-party sources be viewed in a format that can be exploited by the higher DL services (see Figure 1).

This paper describes the Information Mediator Layer in detail by formally describing its role and the functionality of its key services. Moreover, it briefly analyzes some of the existing DLs and DL systems in order to highlight how and to what extent they address the identified mediation functionality.

2 Information Mediator Layer

The *Information Mediator Layer* gives the services in charge of providing *virtual views* of the information space, i. e., a logical grouping or subsets of the information space objects that can be accessed in ways that give benefits over the original form.

The mediators constituting this layer can be classified according to which type of information is addressed by the view they provide, as follows:

- *information organization*: this class of mediators provides two types of virtual views, i.e. the *information object model* and the *virtual collection view*. The former mainly tackles the problem of heterogeneity while the latter deals with problems raised by large volumes of data. The information object model is the view used by the high level services as placeholder and representative for unit of information. It should be powerful enough to capture all the information object characteristics the services need to represent, e.g. multiple versions, heterogeneous object composition, multiple manifestations. The virtual collection view is the "classic" mechanism to organize the information space in multiple sets of objects, each capable to meet a different need. Its functionalities are related to the management of information objects and collections, e.g. creation, access, removal.
- *object manifestation*: this class of mediators provides a *manifestation view*. The manifestation is the way through which the content of an information object is perceived by the user. The functionalities provided by the services of this area are: (i) to access the manifestation while hiding details about its storage, and (ii) the dynamic generation of alternative and more profitable manifestation formats.
- *metadata object manifestation*: this class of mediators provides a *metadata view*. Metadata are data about data and in a DL usually are used to equip an information object with further data other than the manifestation. The functionalities provided by the services of this area are: (i) the presentation of the metadata in a required format, and (ii) the dynamic generation of new metadata. It is worth noting that with the "metadata format" we intend the embodiment used to present them, a.k.a. the metadata schema, while we use "metadata" to intend the manifestation through which the metadata is perceived.

For each of these aspects the Information Mediator Layer has a number of services that implement the corresponding mediation functionality. Some of these services are mandatory in any DL system, others depend on the specific DL application area.

2.1 Information Organization Mediators

These mediators provide and supports the following views: information object model and virtual collection.

The information object model is the common view DL services share about the information object, i.e. the main unit of information managed by a DL. The Information Mediator Layer has a service named *Document Conformer* that, exploiting encoded knowledge about the structure of the documents whose manifestations are maintained by external sources, creates a representation of these documents matching the document model expected by the higher layers services. Current Digital Library Systems supports different document model in order to deal with external content. For instance, Fedora [1] deals with complex objects whose content may be stored internally to the DL as well as reference to external sources. Another example is DoMDL [2] that relies on the mechanism to reference external content. The Content Conformer must be equipped with the logic to configure the mapping among the concrete files and the information object. Examples of such technologies are the various plugins the Greenstone [3] uses.

The virtual collection is the unit of information used to organize information objects into sets. To support multiple views over the same information space and to be able to follow its dynamism, the identification of which objects belong to each collection is regulated via a membership condition, i.e. an expression characterizing the objects. This membership condition can range from the enumeration of the objects identifiers to a logical expression. The Information Mediation Layer has a service named *Collection Virtualizer* that is in charge of linking the collections of documents manipulated by the services with the constituent items physically stored in external archives. This usually relies on the functionality provided by the Document Conformer in order to have an abstract view of information objects. An example of Collection Virtualizer is the Collection Service [4] implemented in OpenDLib that in the context of the CYCLADES project has been used to build collections over an OAI-PMH set of Archives.

2.2 Object Manifestation Mediators

A manifestation view is the embodiment through which the information object content is perceived. The manifestation concept is covered by the information object model and the access is up to the Document Conformer mediator. In this context we have introduced the *Content Transformer*, i.e. a service capable to generate alternative manifestations of a given one. Here, we want to stress the possibility to combine mediators in different workflow in order to enlarge their utilization, e.g. the Content Conformer can be configured to use the Content Transformer in order to dynamically generate a manifestation and use the Content Conformer to have access to the native manifestation. Examples of such technologies are the Fedora *disseminators* and the recent virtual document supported by OpenDLibG, an OpenDLib extension enabling to run jobs on the grid to dynamically generate manifestations as a result of even complex processes.

Due to the importance of the multilingual DLs, an object manifestation mediator named *Content Translator* can be designed to dynamically generate a manifestation in a different language than the original one.

2.3 Metadata Objects Mediators

Metadata is used in different contexts. In this context we mainly deal with descriptive metadata, i.e. the set of data supporting the discovery and interpretation of information objects. It is important to notice that the mediation can go in two directions: the schema used to represent the metadata and the content of the metadata.

An example of the former type of mediation is represented by the *Metadata Schema Mapper*, i.e. a service able to generate alternative metadata representations accordingly to heterogeneous metadata schemas. An example of the second class of mediators is the *Ontology Aligner*, a service able to identify and suggest semantic correspondences between the representational elements of heterogeneous ontologies. Both services rely on a set of mapping rules. These can be injected as configuration parameters or be dynamically derived by the mediator by providing samples of the same metadata when expressed in different schemas, of equivalent terms in different languages, etc.

3 Conclusion

In this paper we have presented the Digital Library Information Mediator Layer, a layer of services providing virtual views on an heterogeneous information space. To comply with the space constraints we only have described the main services, their characteristics and provided examples of concrete implementation in existing Digital Library Systems. An important aspect to be stressed is the customizability of the mediator services. Indeed, they should be reusable in different contexts and thus they have to be configured on the aspects they plan to virtualize.

References

1. Lagoze, C., Payette, S., Shin, E., Wilper, C.: Fedora: An Architecture for Complex Objects and their Relationships. Journal of Digital Libraries, Special Issue on Complex Objects (2005)
2. Candela, L., Castelli, D., Pagano, P., Simi, M.: From Heterogeneous Information Spaces to Virtual Documents. In: Digital Libraries: Implementing Strategies and Sharing Experiences, 8th International Conference on Asian Digital Libraries, ICADL 2005, Bangkok, Thailand, December 12-15, 2005, Proceedings. (2005)
3. Witten, I.H., Bainbridge, D., Boddie, S.J.: Power to the People: End-user Building of Digital Library Collections. In: Proceedings of the first ACM/IEEE-CS joint conference on Digital libraries, ACM Press (2001) 94–103
4. Candela, L., Castelli, D., Pagano, P.: A service for supporting virtual views of large heterogeneous digital libraries. In: Research and Advanced Technology for Digital Libraries, 7th European Conference, ECDL 2003, Trondheim, Norway, August 17-22, 2003, Proceedings. (2003)
5. Tzitzikas, Y., Constantopoulos, P., Spyratos, N.: Mediators over ontology-based information sources. In: WISE (1). (2001) 31–40