OpenDLib: an infrastructure for new generation digital libraries

Donatella Castelli, Pasquale Pagano, Costantino Thanos

ISTI-CNR, Area di Ricerca, Via G. Moruzzi 1, 56124 Pisa, Italy
e-mail: {castelli, pagano, thanos}@iei.pi.cnr.it

Published online: 27 July 2004 – © Springer-Verlag 2004

Abstract. This paper presents OpenDLib, a digital library infrastructure that provides capabilities for new-generation digital libraries. In particular, the paper introduces a document model that can be used to represent a wide variety of document types and describes the open architectural infrastructure that allows for the expansion of the digital library through the dynamic plugin of new services.

Keywords: Digital library infrastructure – Digital document model – Digital library services

1 Introduction

Recent advances in digital library technologies have stimulated new expectations about digital libraries (DLs). It is now recognized that a DL can offer much more than the mere migration of conventional library functions into a digital environment. A DL can manage several types of documents. These documents can have rich structures, be multimedia and multilingual, be annotated, etc. A DL thus has the potential to disseminate information under different expression formats such as seminars, lectures, demos, programs, etc. Moreover, a DL can provide additional services to support the activity of its users. For example, it can provide a cooperative work service for a community of researchers and an authoring service for online education. To support this vision, innovative DL systems have to be designed. OpenDLib [1], developed at the DLlib Center, ISTI-CNR, Pisa, is one such systems.

OpenDLib consists of an architectural infrastructure and a set of core DL services that can be customized to meet the needs of different user communities. The architectural infrastructure provides the necessary software layer for supporting communication and coordination among a set of services that can be dynamically expanded. The core DL services support a new document model, called Document Model for Digital Library (DoMDL), that can represent a wide variety of multiversion documents with heterogeneous content, media, and structure. This rich document model, combined with the extensibility of the set of services, means that OpenDLib can satisfy many of the emerging needs of the DL environments.

In the rest of this brief paper we describe the DoMDL document model and the OpenDLib architectural infrastructure. We conclude with some remarks about future projects.

2 The document model

The main concept of the DoMDL model [2] is the document folder. A DoMDL document folder is a complex modeling object representing different digital documents that are editions, views, and/or manifestations of the same intellectual document.1 By views we mean different perceptions of the content of a document, and by manifestation we mean a physical representation of it.

1
that an XML schema describing each format exists. This mechanism makes it possible for different metadata types for resource discovery, content rating, rights management, administration, etc. to be associated with each edition and each view of a document.

The DoMDL views can be structured in subviews. These may be reference views, i.e., views of other intellectual documents represented by different document folders. For example, DoMDL can represent a thesis on a particular historical period that has been built by assembling a number of texts authored by a student and a number of digitized historical videos and pictures retrieved from cultural archives. This thesis can be modelled as a document folder with a multimedia view that is the aggregation of a set of subviews: there could be a subview for each of the parts authored by the student and a reference view for each of the reused videos and pictures. Each reference view points to the document folder of the corresponding original historical document.

By combining the possibility of moving along the temporal dimension given by the edition mechanism with the flexibility given by the structured and reference views and the possibility of having multiple metadata descriptions for both editions and views, the DoMDL document model provides a very powerful representation mechanism that can be used to handle and disseminate a wide variety of new expression formats.

3 The architectural infrastructure

The OpenDLib infrastructure has been appropriately designed to render a system open, i.e., to permit the expansion and the customization of a system’s functionality. The overall functionality of the system is partitioned into a set of well-defined cooperating services. A new functionality can be added dynamically by including a new service. The new service can exploit the functions implemented by the other services. The OpenDLib infrastructure establishes the rules and provides the necessary “glue” to make this dynamic expansion feasible.

The communication protocol is one of the elements of this infrastructure. It comprises a set of requests that are service-specific, i.e., that depend on the functions exposed by each service, and a set of requests implemented by every service. The latter type of requests ask for general information about the configuration of the service, its current status, the modalities for accessing its functions, etc.

The OpenDLib infrastructure permits the distribution/replication of the DL services on different nodes. This means that there can be multiple instances of the same service in an OpenDLib DL and that, as a consequence, the same functionality can be implemented by alternative groups of service instances allocated to different nodes. Optimization criteria, such as, for example, expected response time or work load balance, can be used to dynamically select the best of these possible alternatives.

OpenDLib also supports highly dynamic architectural configurations: the nodes, service types, number of distributed components, their allocation to nodes, and communication routes can all be modified without interrupting the overall system operation. This dynamic behavior is controlled by a special service called the Manager Service. This service maintains a continuously updated map of the federation and its status, i.e., it keeps track of the available nodes, of their workload and availability, of the services and their allocation to the nodes, etc. By exploiting this information, the Manager is able to check the consistency of the service configuration and can calculate the optimal communication routes among the service components. When a new service instance is added or when a node fails, the Manager dynamically recalculates the new optimal configuration. Periodically, each service component requests from the Manager the best route for its communication and specific information about the status of the federation. This information is used by the services to dynamically coordinate their behavior.

The OpenDLib infrastructure comprises other basic services that implement general infrastructure support functions. The Registry Service is one of these services. This service ensures that only authorized users can obtain the appropriate access, that the users are authenticated, and that both the information they are accessing and their personal credentials are preserved against illegal access. Another service of the infrastructure is the Collection Service. This is a particular type of information space mediator. It supports the dynamic creation of virtual collections. These can be created by either an automatic or human administrator to respond to the needs of emerging communities. This service allows a new community to quickly establish its own dedicated communication means (i.e., something analogous to a specialized journal or to an e-print archive) and to eliminate it in a flexible way if things do not work out.

The current release of OpenDLib implements a core set of application services that offer the functions of a conventional library on DoMDL digital documents, i.e., acquisition, storage and preservation, search, browse and retrieval, and selection and dissemination of documents. Each of these services has been designed to be customizable along a number of dimensions. For example, the services that process user queries can support different types of search operations, e.g., monolingual, cross-lingual, relevant feedback, etc., and different query languages; the storage service is able to support different organizations of the local information space; the browse service can work with different browsing criteria; and so on. The possibility of customizing the service configuration, combined with the openness of the infrastructure and the powerful document model, is a key feature that makes OpenDLib capable of covering a wide variety of different user needs.