Search Strategies for Finding Annotations and Annotated Documents: The FAST Service

Maristella Agosti and Nicola Ferro

Department of Information Engineering – University of Padua
Via Gradenigo, 6/b – 35131 Padova – Italy
{agosti, ferro}@dei.unipd.it

Abstract. This paper discusses two kinds of search strategies supported by the *Flexible Annotation Service Tool (FAST)*, an annotation service that can be used by different *Digital Library Management Systems (DLMSs)*. The first strategy concerns the search and retrieval of annotations, considered as stand-alone documents; while, the second one regards how to exploit annotations in order to search and retrieve annotated documents which are relevant for a user query. This paper describes the proposed search strategies in the light of the architectural design choices needed to support them.

1 Introduction

As observed by [10, p. 274], "the progress of the *Digital Library (DL)* field can be evaluated along several dimensions", one of which is called the *service dimension*, which "characterizes the complexity of processing that DLs and federations of DLs can manage on behalf of clients". In particular, we are interested in studying and developing a service able to add annotation capabilities on the documents managed by a *Digital Library Management System (DLMS)*, so that the service encapsulates all the complex processing needed to provide advanced annotation functionalities and can be easily "plugged" into different DLMSs.

We have designed and we are developing an annotation service for DLMSs, which is called *Flexible Annotation Service Tool (FAST)* [1, 2, 3, 4]. FAST offers basic annotation management functionalities and provides users with advanced search capabilities for retrieving both annotations and annotated documents on the basis of their annotations. This paper will introduce the search strategies supported by FAST in order to search for both annotations and annotated documents and it will describe the architecture of FAST with a particular focus on the architectural design choices which impact the search functionalities offered by the service.

The paper is organized as follows: Section 2 discusses the use of annotations in the context of DLMSs; Section 3 provides an overview of the FAST service; Section 4 describes the search strategies supported by FAST; Section 5 discusses the architecture of the system and its consequences on the offered search strategies; finally Section 6 draws some conclusions and provides an outlook of the future research work.

H.Legind Larsen et al. (Eds.): FQAS 2006, LNAI 4027, pp. 270–281, 2006.

2 DLMSs and Annotations

DLMS are currently in a state of evolution: today they are simply places where information resources can be stored and made available, whereas for tomorrow they will become an integrated part of the way the user works. For example, instead of simply downloading a paper and then working on a printed version, a user will be able to work directly with the paper by means of the tools provided by the DLMS and share their work with colleagues. This way, the user's intellectual work and the information resources provided by the DLMS can be merged together in order to constitute a single working context. Thus, the DLMS is no longer perceived as something external to the intellectual production process and neither as a mere consulting tool, but instead as an intrinsic and active part of the intellectual production process [2].

Annotations are effective means in order to enable the paradigm of interaction between users and DLMSs envisioned above, since they are very well-established practices and widely used. Annotations are not only a way of explaining and enriching an information resource with personal observations, but also a means of transmitting and sharing ideas in order to improve collaborative work practices. Thus, annotations can be geared not only to the way of working of the individual and to a method of study, but also to a way of doing research, as it happens in the Humanities. Finally, annotations allow users to naturally merge and link personal contents with the information resources provided by the DLMS so that a common context that unifies all of these contents can be created.

With this last respect, documents managed by the DLMS and annotations constitute an hypertext [3, 4], since annotations allow the creation of new relationships among existing objects, by means of links that connect annotations together with existing objects. [9] points out that annotations are one of the activities that form the basis of any collaborative effort and for which hypermedia systems are ideally suited, while [12] considers annotations as a natural way of creating and growing hypertexts that connect information resources by actively engaging users. Moreover, DLMSs do not normally have a hypertext connecting information resources with each other; thus, annotations can turn out to be an effective way of associating a hypertext to a DLMS in order to enable an active and dynamic usage of information resources. This hypertext can span and cross the boundaries of the single DLMS, if users need to interact with the information resources managed by diverse DLMSs [2]. This latter possibility is quite innovative, because offers the means for interconnecting various DLMSs in a personalized way meaningful for the end-user and, as recognized also by [10], is a big challenge for next generation DLMSs.

In this evolving context, it becomes crucial to design and develop services able to provide annotation functionalities to many different DLMSs. Moreover, besides offering various annotation management facilities, these services should pay particular attention to offer support for integrating annotations into the information access and retrieval process. Indeed, the possibilities of collaboration and active involvement with digital resources uncovered by bringing annotations into DLMSs require that annotation are an integral part of the way in which