Information Access
Through Digital Library Systems

Maristella Agosti

Department of Information Engineering, University of Padua, Italy
`agosti@dei.unipd.it`

Abstract. The talk presents an interpretation of the evolution of the events and trends in the information access area. Focusing mainly on the last twenty years, particular attention is payed to the digital library system which needs to be envisaged and designed to support the end user in accessing relevant and interesting documents.

1 Information Access

The term *information access* identifies the activities that a person – the *user* – has to conduct to choose, from a collection of documents, those that can be of interest to him to satisfy a specific and contingent information need. The three main actors and aspects that information access needs to address are: 1) user, 2) collection of documents, and 3) access, that is a function or model used in retrieving and accessing documents.

Where the user is the central actor of the situation, the collection of documents is the source from which documents can be extracted in the hope they are of interest, and access is the function that transforms a user information need into a set of documents that are supposed to satisfy the user information need.

When the collection of documents grows and reaches a size that makes a manual inspection of the documents prohibitive, the construction and management of the collection together with the application of the access function are managed in an automatic way through a digital library system. In parallel to the growth of the size of the collection, many other changes happened both on the size of the individual documents and on the type of the documents which form the collection. Now the size of a document can span from that of a textual abstract of just few hundred words to that of a complete book or a video where multimedia representations need to be managed and accessed by the final user. So the final user is confronted with a new and more complex context of work, where the diversity of the types of documents can increase together with the rise of the size of individual documents.

This means that the context of reference in building and using a digital library system changes over time requiring the creation of new architectures and approaches. The talk critically analyzes the evolution of the modeling of the information access systems, mainly in the last twenty years, relating the general analysis to conducted experiences[1].

---

[1] M. Agosti (Ed.). *Information Access through Search Engines and Digital Libraries.* The Information Retrieval Series, Vol. 22, Springer, Heidelberg, Germany, 2008.

2   Evolution of Information Access Systems

Early Days. In the early days of computer science, the common approach was to consider the specific type of documents constituting the collection and to manage and design the system and applications around it. The attention of the system designer was concentrated on the specific type of documents, mostly because the available technology limited the possibilities of representation and management only to textual documents.

1977–1986: The Last Decade of Centralized Systems. The attention of the system designer was prevalently focused on the textual collection of documents more than on the user and the specific function or model to implement in the system. Still in this decade systems were able only to manage a single type of documents, and the systems were named in accordance with the specific type of document collections they were designed to manage in a specialized way. It was in those years that a new generation of library automation systems started to be designed with the purpose of enabling the managing and accessing of a combination of different types of data: structured catalogue data together with unstructured data representing the contents of the documents. This new type of library automation systems, that can be considered the "ancestors" of present days *digital library systems*, also supported the interactive retrieval of information.

1987–1996: Towards a User-Oriented Decentralized Environment. At the beginning of the decade, researchers took new directions, trying to support the users with new types of access models or with a combination of different models. New approaches to information access were proposed. Some of those proposals were based on the hypertext paradigm, and those ones made use of the links that exist among documents and descriptive objects. Some successful approaches used a two-level architecture to represent on one hand the collection of content objects, and on the other the content representation structure.

1997–2006: The Overwhelming Amount of Digital Documents. During this decade it became clear that it was necessary to face the continuous growth of diverse collections of documents in digital form, so major efforts faced different aspects related to the growing of diversified digital collections. Multimedia access is the other relevant area that researchers started to face in a systematic way and for different media during the decade. The complexity of the management of collections of multimedia digital documents can be faced in particular for information access purposes, but also from a general architectural point of view, that is, the area of *service oriented* digital libraries and digital library systems.

From 2007 On. The new decade that starts this year is opening up new exciting challenges, both on the side of architectures and on the contents for the construction of a new generation of digital library systems.