Libri, 1999, vol. 49, pp. 11–15
Printed in Germany · All rights reserved

Libri
ISSN 0024-2667

The Role of Libraries and Librarians in Organising Digital Information

PETER INGWERSEN
Royal School of Library and Information Science, Copenhagen, Denmark

This article analyses the role of libraries and information professionals in a digital library environment. The major focus for discussion is the availability of improved intellectual access to knowledge sources in a digitised world. The author argues that purposeful knowledge organisation, information filtering, and management skills are mandatory *a priori* elements that support accessibility and constitute the future roles of information specialists and services. It is argued that the complexity of the digital library infrastructure determines the degree of human involvement. A three-level digital library infrastructure scenario of increasing complexity serves as the framework for the discussion: the stand-alone, the distributed, and the integrated digital library.

Introduction

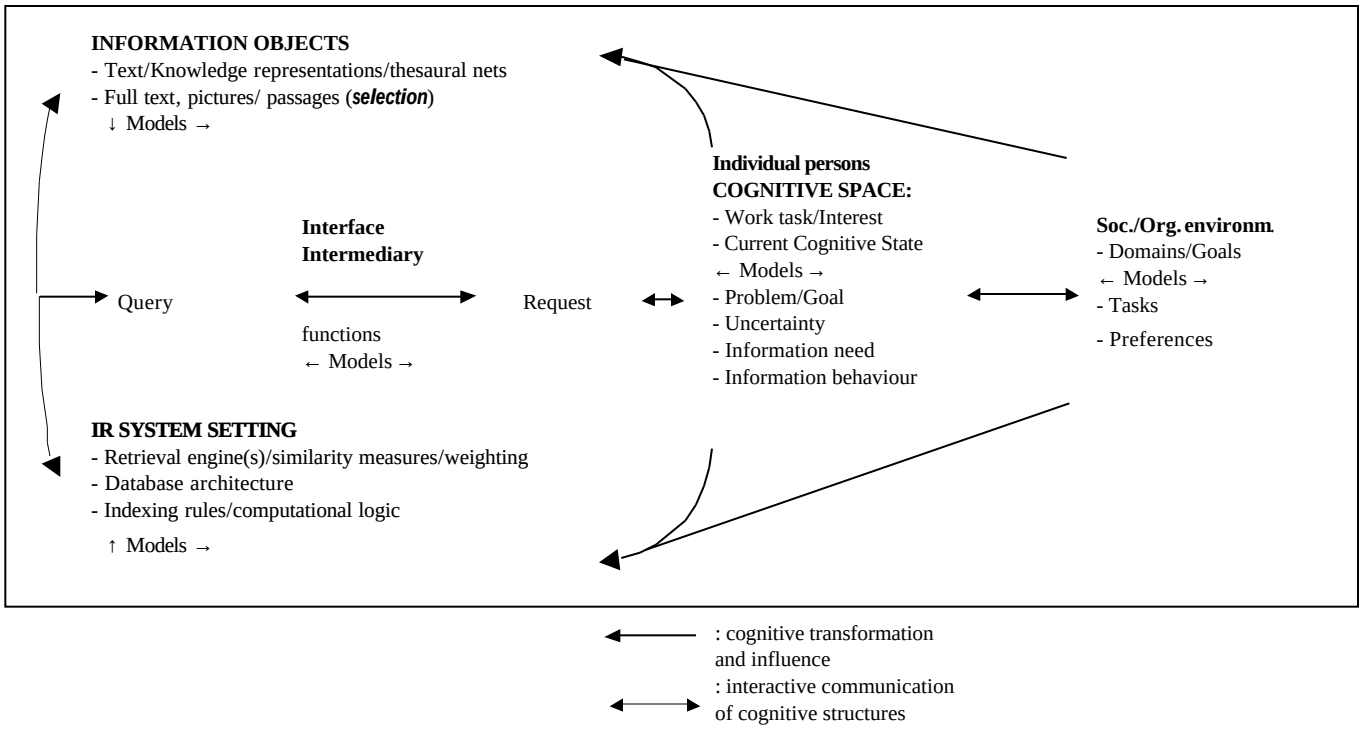
In an Anglo-American-Scandinavian context, the public, research and university libraries have positioned themselves as the main vehicles for the organisation, access, distribution and use of stored knowledge of quality. The role of these institutions and associated information services, such as online providers of various kinds, is to contribute to the education, leisure, and thus to the socialisation processes of our societies. In other parts of Europe and the World, libraries take up a weaker position by being dedicated to specific professional groups, domains, or knowledge sources. In this context the World Wide Web is regarded as an unstructured blend of more or less reliable sources of different information quality – foremost an attractive marketing tool in colours. With the digital library concept in mind this situation may be more balanced in the future and favour global interests.

To the information professional, the librarian or the information provider, the role has similarly been actively to develop and maintain the management of a wide range of structured and organised knowledge resources by providing bibliographic and physical access, but not necessarily *intellectual access* to such resources (Wormell 1992). The latter access mode is partly within our grasp due to the digitisation. However, a digital library is not simply a "world of electronic objects in cyberspace." Paraphrasing Manfred Kochen's words from 1974: it is a "growing encyclopaedic system for summarising and organising knowledge." Like all archives, museums and libraries, the digital library presupposes *organisation* of its holdings, *managed* according to a *purpose*. Without goals and organisation, any access and utility become random or even unreliable.

In the digital library context, the role of the librarian is primarily to manage the organisation of such collections: digital librarianship. Clearly, the complexity of the *digital library infrastructure* determines the human involvement. We may operate with three digital library scenarios of increasing infrastructure complexity: the stand-alone digital library collection, the distributed digital library, and the integrated digital library approach.

This article has sections on the basic components, interactive activities, and human factors

Dr. Peter Ingwersen is Head of the Department of Information Retrieval Theory, Royal School of Library and Information Science, Birketinget 6, DK 2300 Copenhagen, Denmark. Email: pi@db.dk
This article is based on an invited lecture at the Digital Libraries Asia 98 Conference, Singapore, March 17–20, 1998.



that should be taken into account when design-

Figure 1: Cognitive communication model of information interaction in information science. The notion of 'Individual persons' covers both users *and* librarians' cognitive spaces during IR and selective acquisition (Ingwersen 1996: 9).

ing digital library systems to fit human knowledge requirements. A three-level digital library infrastructure is outlined, followed by a discussion of the access possibilities in traditional services and digital libraries leading to the roles and activities to be foreseen for information specialists and services.

Information interaction

The model shown in Figure 1 is of general nature and points to the central interactivity, components, and characteristics of (multimedia) information systems (Ingwersen 1996). A digital library can be regarded as such a system.

Commonly, the design of information retrieval systems has been done by focusing on the left-hand side of the model displayed in Figure 1. This very traditional approach constitutes the system-driven or *algorithmic approach*, for example also applied to IR experiments. This design mode is similar to that currently preferred in the case of digital library designs, i.e. quite technology-driven. Observe, for instance, the proportion of papers contributed to digital library conferences globally. The technology-dependent issues are paramount (ACM 1997). Naturally, all information systems consist of a System Setting that interacts with the Information Objects selected by the system. An interface facility ensures that a user may interact with the objects and/or setting, e.g. by providing relevance feedback so that the search engine(s) may find alternative objects based on user relevance judgements. First during the 1990s, designers of IR systems and test scenarios have recognised the need for involving the Cognitive space of the user and the Social/Environmental dimensions, Figure 1, the right-hand side (Ingwersen 1996). Only very recently best match retrieval models started to take into account the *dynamic nature* of a user's information situation, his goals, tasks, cognitive state and behaviour (Campbell and van Rijsbergen 1996). The usual assumption behind IR experiments is that the cognitive space is static during search sessions.

Therefore, what seems important to the digital library scenario is the degree of involvement of the user and use dimension. Regardless of the nature of a digital library, the information envi-

ronment should influence its design, and information specialists and scientists seem quite suitable for this enterprise. The problem is, as we see it, that libraries and librarians have commonly in the past based their organisation of and the access to library collections on the document contents – not on the context.

With Buckland's concepts of *information* (1991), we may observe that information is seen as a "thing" – not as a "process" leading to "information as knowledge". This is perhaps understandable when talking about large national library collections. But in the case of dedicated domain-dependent libraries this closed view is less obvious. In fact, the online bibliographic database services have since the 1970s been far more user-centred in their knowledge organisation and access policies than libraries. The records in the textual online databases are far more enriched (by human efforts) than the records applied to the various attempts to create OPACs (online public access catalogues). Essentially, the drawback of the OPACs was the poor subject descriptions originally made in the libraries. The digital library concept displays a continuation of the OPAC attempts, however with the fortunate characteristic that the Information Objects are electronic and available in their complete form. Evidently, this fact poses the usual problem of quantity versus quality of accessible data, found everywhere that IT produces changes to an existing situation. An ensuing question is consequently whether or not digital library objects should become conceptually structured and enriched, leading to an improved intellectual accessibility?

Finally, one may point to the issues of: 1) *selection criteria*; 2) *alteration* of information objects; and 3) keeping track of *versions* over time in the digital space. On the Internet old pages commonly cease to exist when altered ones take their places. A traditional library can only with difficulty change the contents of its objects and is not supposed to do so for reasons of preservation and custody. Selective qualitative acquisition is the preferred method by which new versions (editions) are regarded as new objects. Similarly, a bibliographic database is believed to have its various means of access and possible full-text objects stored and fixed once they are acquired and indexed. This belief has its root in the technology of paper, but does *not* have to hold in the case of, for instance, distributed digital libraries where there is less central control with acquisition policies being applied by individual members. So, should all versions of an available electronic document be preserved – the archival function – or only the "published" one – the usual library function? And – is the qualitative selection of objects a must? These issues call for a significant role to be played by librarians in collaboration with personal and commercial publishers.

The Digital infrastructure

Following Saracevic (1997) one may distinguish between three digital library scenarios: a narrow digital library collection; a distributed network-based collection; an integrated framework. The first scenario is, in my interpretation, a stand-alone digital library characterised by a fixed location (a server) and with an organised acquisition policy. The information objects, the system setting and the interface are under central management control. It may be tailored to specific knowledge sources of all suitable kinds in a domain and serve a limited user population. Analyses of the information and user environment is thus possible (Figure 1) in order to design the system to fit the human being. *Navigational and filtering skills* are consequently mandatory capacities for the information specialists managing the library and operating as human gateways in order to select and incorporate (buy and own) the appropriate knowledge sources to be acquired. An existing library may carry out this kind of venture, partly based on its own collection. This type of a digital library resembles the guided-tour "manifestations" of virtual museums (Table 1).

The increasing complexity of the infrastructure may take the form of several independent distributed digital libraries. The information objects and retrieval settings are scattered across networks and held together by agreements and alliances – like in today's national and regional OPAC networks. An intermediate form consists of a stand-alone library which provides *digital referral* to sites and collections by means of a single gateway – the Interface (Figure 1) – and by defined managerial and acquisition policies. Its reliability and use then depends on its qualitative

Table 1: Basic access lines to selected information services.

Access lines	Net-worked Access	Biblio-graphic Access	Physical Access	Intellec. access Feature-based	Intellec. access Concep-tual
Trad. library		X +	X		
OPAC	X	X +			
Online Service	X	X ++	full-text dbs.		
WWW	X	X	X	(hyper-links)	hyper-links
Virtual Museums	X	X +		X	Guided Manifes-tations
Digital Library	X	X (++)	X	X	?
Digital Referral	X	X ++	indirectly	(hyper-links)	hyper-links
Distrib-uted	X	X ++	X	X	?
Inte-grated	X	X ++	X	X	?

+ refers to the degree of subject indexing.
() indicates automatic generation of access means.

selection policy, the navigational, *and* the indexing / presentation skills of its information professionals. Although that library – like Yahoo! – may provide enriched access to the various sites, it lacks ownership and control of the collections to which it refers. The physical access is indirect and the intellectual access into the objects referred to depends on the remote hyperlink construct.

The last scenario envisioned by Saracevic (1997) consists of an extension of the distributed digital library into an infrastructure that *integrates* the total complexity as displayed in Figure 1. That is, it is an environment bringing together collections, societies & services, as well as people and publishers in support of the full cycle of creation, dissemination, use, and preservation of recorded knowledge.

The Access issue

Bibliographic access is understood as the entry of surrogates for information objects, e.g. in the form of indexes and catalogues. Descriptive indexing can be carried out automatically; whereas subject indexing commonly implies a degree of human interpretation and thus also loss of information. Physical access entails the access to the information objects' proper. Intellectual access can be divided into two types: 1) feature-related access to objects, e.g. in the form of passages, based directly on their contents or inherent features, such as phrases, citations, figures and captions, image forms, etc.; and 2) conceptual access to the information embedded in the objects based on a knowledge of the actual information situation of the reader / user (i.e., aboutness). Feature-related access implies the application of already existing and yet to be developed best match IR processing tools (Ingwersen and Willett 1995), and is often regarded a substitute for human subject indexing. The conceptual type of access presupposes a qualitative dimension of retrieval tailored to the actual user and the domain context which, thus far, has not been made workable in complex knowledge environments (Smeaton 1992).

The IT revolution during the 1980s, characterised by the introduction of the inexpensive personal computer and electronic networks, has provided a fourth access possibility: *networked access*. But it has also added to the complexity of the information environment and introduced the alternative knowledge distribution channels of handling full text and other media objects. The digital library concept further changes this environment into quite a complex scenario in which instant feature-based *intellectual access* for users is possible (Table 1).

What has changed is the information infrastructure as outlined above. Today, and in the near future, we have five or more major channels or *access lines* concerned with knowledge and entertainment distribution simultaneously associated with library acquisition *and* the dissemination / retrieval / use dimensions of transfer. They are: 1) traditional access to and provision of information *only* by means of *representations* of documents or information objects, i.e. library catalogues and bibliographic databases – the bibliographic access aspect; 2) existing physical access to non-electronic documents – by *being in the library* (or archive / museum) or by getting a copy delivered directly to the user; 3) the physical access to remote electronic documents or events via bibliographic access and unstructured (chaotic)

distributed networks, i.e. like presently on the World Wide Web; 4) the feature-based intellectual access to structured and managed digital library collections; 5) the conceptual intellectual access, primarily made possible by a knowledge of the visitor demands and professional intervention and construction. This last option is today perhaps only available in virtual museums. To the digital library the question is should information professionals in a similar fashion attempt to enrich or conceptualise the intellectual access? And is it feasible?

Conclusion

All access lines have in common the requirement that somebody *manage* the infrastructure by means of processes of *input*; these include: acquisitions policy; data mining and capture by navigation; quality criteria and control; knowledge organisation; and processes of *output*, i.e. communication; visualisation; interactive retrieval, navigation and browsing; filtering; extraction and summarisation; use, work task, and interest fulfilment. Both the existing as well as the digital library, when placed virtually on the Net, may provide such access lines to users *and* to the librarians during their acquisitions process; the challenges are to supervise the named processes as well as the increasingly complex infrastructure in a cost-effective *and* user-centred way. It is my belief that information professionals can best meet the challenges on the *input side* and together with computer scientists, HCI and management researchers can be responsible for the design, evaluation, and maintenance of the output processes, including the economic aspects of the collections. An important additional activity is constituted by the *informetric analyses* of communication patterns that take place via digital library services (Almind and Ingwersen 1997; Ingwersen 1998). The tasks of digital librarianship entail improved professional competencies in which scientific, research methodological, managerial and economic skills are integrated with communicative, navigational, information seeking, retrieval and analytic design knowledge (Wormell 1996; Ingwersen 1994).

Essentially, the digital library poses similar challenges to the profession and to institutions, as has been the case in the past and present for the physical library: management by integration and knowledge of users. The solutions to the novel infrastructures of digital libraries can not be made of technology alone but must be multi-dimensional and interdisciplinary within the human context.

References

ACM International Conference on Digital Libraries. 1997. Proceedings of the 2nd ACM International Conference on Digital Libraries, edited by R. Allen and E. Rasmussen. New York, ACM.

Almind, TC. and P. Ingwersen. 1997. Informetric Analyses on the World Wide Web: Methodological Approaches to Webometrics. *Journal of Documentation* 53(4): 404–26.

Buckland, MK. 1991. Information as Thing. *Journal of American Society for Information Science* 42(5): 351–60.

Campbell, I. and C. van Rijsbergen. 1996. The Ostensive Model of Developing Information Needs. In: Information Science: Integration in Perspective edited by P. Ingwersen & N.O. Pors. Copenhagen: The Royal School of Library and Information Science, 251–68. ISBN: 87-7415-260-2.

Ingwersen, P. 1994. The Human Approach to Information Science and Management: The Framework and Prospects underlying the new Danish MSc Programme. *Journal of Information Science* 20(3): 197–208.

Ingwersen, P. and P. Willett. 1995. An Introduction to Algorithmic and Cognitive Approaches for Information Retrieval. *Libri* 45: 160–77.

Ingwersen, P. 1996. Cognitive Perspectives of Information Retrieval Interaction: Elements of a Cognitive IR Theory. *Journal of Documentation*, 51(1): 3–50.

Ingwersen, P. 1998. The Calculation of Web Impact Factors. *Journal of Documentation* 54(2): 236–43.

Kochen, M. 1974. Principles of Information Retrieval. Los Angeles: Melville Publishing Company.

Saracevic, T. 1997. Digital Libraries: Interdisciplinary conceptions, challenges, opportunities. New Brunswick, NJ: School of Communication, Information and Library Studies, Rutgers University. (Handout: Honoring the Scholarly Communication Center; personal communication).

Smeaton, A. 1992. Progress in the Application of Natural Language Processing in Information Retrieval Tasks. *Computer Journal* 5: 268–78.

Wormell, I. 1992. Understanding Information. Copenhagen: Royal School of Library and Information Science, p. 22–27. ISBN: 87-7415-237-8.