Document architecture draws a circle: on document architecture and its relation to library and information science education and research

Mats Dahlström and **Mikael Gunnarsson**
Högskolan i Borås
Borås, Sweden

Abstract

The architectural metaphor used in analyzing compound meta-objects such as referential databases might be applied also to the primary records these referential tools point to, thereby making way for the study of document architecture (DA). Library and information science has by and large been focussing on possible ways of determining the meanings of the objects of "input", whereas the materiality and the textual structure of the objects have been regarded as transparent, offering little or no room for problematization and discussion. The article argues for a revaluation of this somewhat delimiting perspective. Digital production and distribution reframe the ways in which objects and meta-objects might be construed. The mismatch of traditional library institutions and systems (where the printed codex book and its derivatives have been the standard of measurement) and digital carriers for bodies of text and the different architectures of these, suggests our great need for new fields of LIS research, where DA might prove a valuable tool. DA studies might also be useful in re-theorizing traditional reading and writing technologies and their conditioning of textual carriers.

Printer-friendly version

Introduction

Throughout the ages, one of the main interests for the librarian and for library and information science (LIS) has been different types of referential and bibliographic databases. These (including catalogues and indexing & abstracting services of different kinds) have been studied as compound meta-objects (1) providing pointers to and descriptions of other objects of primary interest to the end-user. Characteristics of these meta-objects have been critical factors in determining the values of their use. Some of these characteristics may be metaphorically described as belonging to the *architecture* of meta-objects, which in practice is usually expressed in the "hierarchical" structuring of databases into records, fields and subfields or the "relational" structuring of databases into entities and relations.

In no respect has the same attention been given to the kind of objects - the textual containers - that they point to, as to their characteristics other than those relevant in predefined cataloguing practice and, in the case of citation indices, other than those that link scientific works together. Moreover, in Knowledge Organization, the preoccupation is with the rules and methods for classification and indexing. It seems as though one is reducing the objects of "input" to their mere possible inherent meanings. The basic material form and the factual, as well as the more abstract, textual structure of the objects have, to some degree, been regarded as transparent and self-evident, almost axiomatic starting points, offering little or no room for problematization and discussion.

We argue for a revaluation of this somewhat delimiting perspective, which we see as a *per se* reasonable outcome of

the multicentennial hegemony of print technology. As print technology is complemented with digital technology, and as the latter is on its way to reframe the ways in which meta-objects might be construed, we want to emphasize the practice of *document studies*, hitherto largely neglected in LIS education and research. These studies must be conducted free from media chauvinism (2), and may be realized in such subfields as e.g. production, management, delivery, organization, history, theory, materialities and sociology of documents as well as, of prime interest to this particular article, *document architecture* (DA). In this respect we may also relate to a number of disciplines - e.g. literary theory and analytical bibliography - where basic assumptions have been questioned pertaining to that supposed basic atomary unit in the architecture of large document corpora, notably the codex book or a journal issue. Digital carriers for bodies of text are handled by traditional, print-based institutions only with great difficulty, at times with considerable loss of information, and quite often in a misleading way. A brief consideration of digital documents vs systems such as legal deposits or descriptive cataloguing might by itself suggest our great need for new fields of research, education and course syllabi in LIS schools. (3) Studies of digital production and distribution might also prove to be of value in re-theorizing technologies for reading and writing and their conditioning of textual carriers. This crossover may yield a reassessment of document architecture in its own right, and might thus make way for a rebirth of a subdiscipline largely disregarded in LIS. But let us first dig somewhat deeper into the very concept of DA.

The notion of document architecture

With document architecture (DA) we mean several things implied by the metaphor *architecture*, besides the formal definition of ISO 8879:1986 which reads "Rules for the formulation of text processing applications" (ISO 8879:1986 clause 4.97). We believe that as the architecture of a building discloses a lot of e.g. the architect, his skill, the architectural style and its underlying truth-holdings of a functional and aesthetic building, DA discloses a lot of the practices and underlying theory of the production of the document.

DA as a concept is closely connected to a certain kind of "text processing applications", general markup languages (GML), where content is separated from presentation.

This structural strategy for digital representation is contrary to strategies, where documents are represented and processed with respect to their outlook: their positioning on the screen or on a sheet of paper. The latter perspective makes no difference between a picture and a sequence of alphanumeric characters. The former takes account for and stores unambiguous information on a document's structural elements, and is exemplified by SGML, XML and originally even HTML. (4) This strategy may also contradict traditional text production strategies, where the writer tends to focus on appearance. Our observations when teaching LIS point to the fact that students learning HTML have extreme difficulties in understanding this concept of separation.(5)

In the context of general markup languages the concept of DA serves the purpose of signifying every more or less compelling system of rules that "deals with the document's purpose, its audience, its possible media, and the variety of ways that the document will be played, performed, displayed, accessed, transmitted, or read". (Turner, 1996: 10) Thus a bundle of instructions on how to render a document on a screen (such as a style sheet), as well as how packets of metadata may be inserted into the document, is part of the DA.

However, as all these rules depend on more or less temporary conceptions of how we might adequately produce documents and on technological constrains, the study of DA has a much broader scope than the purely technical bits.

The metaphor may be said to apply *both* to how these ideal models of text production work and become expressed in different media, both digital and analog, as structures of different document types, *and* how these models (or styles) form taxonomies of different but related styles. DA may then be studied as a matter of how to build, and - from the perspective of aesthetics - as investigations in different styles for different epochs, cultures and genres. And perhaps in the context of a "sociology of documents", as John Seely Brown and Paul Duguid seem to suggest:

> "To fully assess the document's evolving role requires a broad understanding of both old and new documents... They are also a powerful resource for constructing and negotiating social space." (Brown & Duguid, 1996)

With this conceptual background of DA, let us now turn to three contexts, where DA might be a tool for fuller

understanding: information seeking, document (re)production, and theories of technology.

Document architecture and its importance for Information seeking

Information seeking is often described as a "constructive process characterized by uncertainty and confusion", where an "information search is a learning process".(Kuhlthau, 1993, pp 8-9) Thus the description seems to advocate for a perspective on research and education based on user perspectives, leaving out the system-oriented perspectives of the so called "bibliographic paradigm" (ibid., p. 1). There is no denying the fact that information seeking is always performed as actions highly dependent of personally situated and contextual factors, but it is our view that if the impact of a changing technology on information seeking will be determined, this must include an analysis of the relationships between DA and document representation - an analysis that calls for questions usually attributed to the domains of indexing and cataloguing, but extends their scope radically.

If we accept that analyses of information seeking missions sometimes may be characterized as investigations into how articulations of the needs of the users relate to the way documents are represented in different types of meta-objects, then we are also allowed to study the interplay of DA, document representation and user in an information seeking context. As far as we know, this kind of study has hitherto not been dealt with explicitly to any satisfying degree in LIS domains. This is not the forum to elaborate further upon this, merely to point out how different DAs are comprehended by different actors, and to hope for increased interest in these questions. Further empirical investigations into these matters would be welcomed, as we base our presentation here solely on our acquaintances with the web and our observations of how students deal with it.

Examples of how architectural changes affects information seeking

In learning to use the web, students in general have to get rid of several misunderstandings of the underlying technologies. Pollock & Hockley (1997), in a 1995 study on PC-literate users conducting searches on the web, came to the conclusion that "users need at least some understanding of basic Internet concepts in order to carry out successful searches". This study shows, among other things, how users fail to understand that articulation of their needs must be formulated according to how technology makes up a search language. Among LIS students the awareness of this is increasing, but the situation is still too much alikened to bibliographical database searching. There seems to be a strong tendency towards being as precise as possible in the hope that search engines will present *final* answers. It certainly would be better if search services expressed "search results as 'suggestions' rather than 'hits'" (ibid.).

The misconceptions of how web searches may be performed are fairly natural as bibliographic systems always have been important for the LIS field and are emphasized in most courses on information seeking, where the main question at hand seems to be to find relevant textual objects and discriminate irrelevant ones.

In the following we will emphasize some architectural conditions of web documents that affect the use of search engines and other search services on the web. An example will be thoroughly elaborated, viz. how the concept of the title of a document plays an important role in a known-item search.

Different worlds - different titles

A known-item search occurs when the user has a limited but correct description of an existing document. The user is sure of the fact that the document exists, that its title and author are explicitly stated somewhere in the document, and these assumptions are true to the actual state of the docuverse. This situation is fairly different from when a user searches for unknown documents.

A known-item search in a bibliographic database is highly supported by the predictable, traditional DA. Almost every book has a stated title, responsibility, editor etc. Different codes for document description, as for instance the AACR2, prescribe that a representation of a document must include a statement of a document's title and author. It is obvious that, to some degree, conventions for producing documents have been adapted to facilitate such descriptions for the producer. The CIP is the most obvious example, where a bibliographic description is inserted into the document in a prescribed position. Thus documents making use of the CIP is an instance of a document type that defines a position as well as a form of expression for metadata (6) conforming to library practice. These metadata are extracted, in accordance with cataloguing rules, from other parts of the document, e.g. the title page,

which in its turn is another element of the conventional DA of a book.

Consequently, the user searching for a known book may rely on the fact that what he or she thinks is the title of a book is also true with respect to its representation in bibliographic databases. Exceptions are of course not too rare, due to orthographic fallacies and the like, but there are apparent correspondences between rules for generation of document representations in bibliographic databases and common sense apprehensions.

The user carrying out a search on the web, in Alta Vista, Infoseek or some other "search engine" can not rely on such a correspondence - for several reasons. For example, according to Gudivada et al. (1997), it has been estimated that 20 % of the objects on the web lack the title tag.

Admittedly, formal HTML specifications do prescribe the declaration of a title, but since current web browsers seem more or less to ignore if the HTML object conforms to formal rules or not, and since the producer of a web document not always recognizes the importance of significant titles, the result is that many web documents anyhow lack the title or have titles that make no sense. The problem may become less frequent with a transition from HTML to XML, as the latter imposes more strict regulations on the producers of browsers (XML processors). This may happen if XML processors, as is suggested (Goldfarb, 1998), won't accept *ill-formed* documents, which means that an XML object, nonconformly encoded, will not be rendered at all and consequently will be unusable. With respect to this possible evolution it may be said that XML architecture is imposing further constrains on the writer and seems limiting, but at the same time makes expressions of this architecture more predictable, for generators of document representations as well as for readers. In the meantime it is a fact that the user of a search engine, especially a LIS student, expects that a search term attributed by `title:`(7) will match all documents with that term in the title. The fact is though that as the (formal) title is that part of the text manifested in the upper bar of the window, both readers and writers tend to regard the first heading or another visually prominent part as the title. Neglecting the meaning of the title tag may also cause the writer to omit it. The mismatch of the architectural definition of a title and the user's is obvious, and may be of critical importance, especially when the writer has represented the title with an in-line image, as images aren't indexed at all by search engines.

As cataloguing codes are beginning to adapt to new media in order to incorporate pointers to resources on the web in library catalogues, rules for determining the title of web documents are formulated. This is as far as we can see not done in a way that takes account of already existing (technical) definitions of the title or the unique architectures of web documents. In fact the ALA in a report last spring (CC:DA 1998) came to the conclusion that existing schemes for metadata (TEI and the Dublin Core) are of limited use in catalogue integration. It all seems as a tentative strategy to treat web documents as expressions of the (inherently different) architecture of journal articles, books or other types of printed documents. This strategy often concludes that the title is the visually most prominent part of the web object's rendition, usually a declared heading or the text of a GIF or JPEG image, in analogy with AACR2 rules for book description, which (naturally) makes references to visual appearance.

This may lead to a situation where not only do search engines and users diverge in their definitions of the title, in their models of what constitutes the architecture of a web site, but also where cataloguing rules develop another conception of the DA of web objects.

So if we say that different worlds take account of different titles, we may also say this about other architectural phenomena.

In a recent article on "citation in the digital era", Mats G. Lindquist remarks that "libraries are tied to the book and consider journals to be fragmented books". (Lindquist, 1999) A statement that may be compared to the remark of Luciano Canfora, that "[f]or librarians, the scroll was the 'unit of measurement'", implying that estimations of the contents of the Alexandrian library in most sources were highly exaggerated, since they were derived from "the practice of counting not works but scrolls". (Canfora, 1990, p. 189) Since 1994 we have seen that users may open *files*, *documents*, *URLs* and *sites* in the Archive menu of the browsers. At the moment we seem to be opening *pages* (in Netscape), though the word *page* here merely resembles what is usually meant by that word. Abundant examples witness to the fact that, as metaphors for a new technology are derived from older information technologies, this may cause unnecessary ambiguity (8). The concept of document is in itself somewhat of a metaphor, whereas it signifies a way of establishing borders for a coherent whole, by the help of reifying a conception of a collection of entities into a coherent work, a document.

Consequently, we believe that all this points to the relevance of a DA perspective, as it is important to elucidate the

unique nature of the outcomes of a modern technology.

If we understand *technology* in its broadest sense, not restricted to *modern* technology, we may approach its essence in much the same way as Heidegger in *The Question Concerning Technology* (1977), where technology is described by traditional "instrumental" and "anthropological" definitions. Technology, then, may be seen as "a means to an end" where "human activity" takes place in "bringing-forth". Thus writing and reading may be seen as activities for bringing forth human expressions - thereby using tools, which put necessary constrains on the activities. In these activities there is no way of becoming independent of technology. Language as well as inscribing tools (like the pencil or the keyboard) and storage media form a technology for writing and reading. A technology in which we "dwell", rather than use as a simple tool.

Thus, if we accept this view of technology as something that transforms, as well as is transformed by, mankind, any change in technology will affect the possibilities of expression, and this must be regarded as a critical factor in analyzing information seeking.

Document architecture as a tool for understanding the production and construction of literary works

This view of technology might be of some assistance to us when considering the importance of DA in the study and theory of *markup languages* as tools that shape, as well as are shaped by, our conception of possible DAs. As noted above, the term DA even originates in the realm of markup practices.

Markup as a theory of DA

DA becomes important when trying to understand metagrammars or schemes of how textual works are and can be constructed, most apparently so when we deal with the aforementioned markup languages for either the direct production of new digital works or the digitization of works previously represented in paper documents. As Michael Sperberg-McQueen notes (1991), a markup of a text (9) is in fact a theory, or a theoretical statement, of this text. Extending this view, then, a general markup language is a general theory of texts. Note that that this usage of "texts" is more affiliated to that of text sociologists such as e.g. Donald McKenzie and Jerome McGann than to that of the kind of bibliography represented by e.g. Thomas G. Tanselle. The text concept of the latter is limited to the sequence(s) of linguistic alphanumerical signs, whereas the former aims to include, in McGann's terms (McGann, 1991), linguistic signs *as well as* bibliographical codes (such as typography, nature and quality of the storage medium, colouring etc), properties traditional bibliography would describe as belonging to the document level rather than to the text level. A number of tags used in various DTD:s in fact deal with the *architecture* of documents, since a markup language is a tool for structurally handling aspects and elements on different hierarchical and conceptual levels of the document, not only textual strings at the linguistic level. In this way we might consider certain families of markup languages as theories of particular types of documents or of genres, and as statements as to how these documents or groups of documents can in fact be constructed at all, and, finally and importantly, as tools perhaps shaping the very production of documents and their architectures.

Consider the universally intended grammar of SGML. SGML is not a markup language *per se*, but might instead be regarded as a markup *metagrammar* (certainly when we treat SGML in terms of an international standard, i.e. ISO 8879:1986), a system of rules out of which actual, derivative markup languages in turn might be constructed. In practice, user communities collaborate and discuss what types of elements are needed for certain types of documents and their markup (in particular when digitizing older literary works), and a formal Document Type Definition (DTD) is agreed upon (10). This process has to include the troublesome definition and consequent agreement on what are the important inherent elements in the documents at hand, i.e. an agreement on what *constitutes* the documents. (11) This DTD has to be followed in order for documents to be universally exchangeable and processable within the intended user group. SGML was from its beginnings bound to the notion that DAs in general were characterized by ordered hierarchies of content objects, textual blocks that could not overlap. This was, in a manner of speaking, the nature of documents according to this theoretical view. Attempts were even made to establish theories of the ontological status of texts as consisting of content objects in a strict hierarchical order (DeRose et al, 1990 ; summed up and elaborated in Renear, 1997). This underlying philosophy of SGML defined and constrained the development of derivative applicational grammars - markup *languages* - such as TEI and HTML, thereby restricting these tools to a particular concept of DA. It soon turned out, however, that there in fact

existed a number of document types that did contain overlapping hierarchies (12), contrary to the general theory of SGML, and therefore could be marked up according to a DTD of SGML only with considerable loss of adequacy. Where does that leave the "general" in General Markup Languages?

DA draws a media transitional circle

The explosive growth of the web and the consequent widespread use of HTML and its tools, along with more or less explicit demands of adequately encoded HTML documents, have resulted in equal growth and spread of this architectural conception of SGML. The architectural view, upon which SGML and HTML originally were based, thereby might justify itself in a, as it were, roundabout way. The technological tool, the thing with which to write, will affect the thing written. The particular DA conception in SGML and derived languages will affect the production, collection, definition and analysis of documents, both as documents are originally produced electronically, and as paper-bound works are being digitized, and might consequently be of major interest to LIS. (Interestingly enough, digitization has offered opportunities for reframing the contents and architectures of manuscripts and printed books, forcing the digitizing agents to express, code and group digitized reproductions of not only the documents as entities, but also *parts of* the documents, whereby the entities as well as the parts can be handled in numerous ways, e.g. restructuring, searching, downloading, printing (again!) and commercializing (The Center for Retrospective Digitization, 1999). This makes way, in a manner of speaking, for a more literal version of the infamous *deconstruction* of works, but this time a deconstruction not of the inherent meanings in works, but of the direct textual expression as manifested in the physical documents.)

With this in mind, we might begin to analyze in what way text application tools such as word processing software or markup languages condition how we are able to produce and reproduce (when digitizing) works, and in what way they constitute statements on possible DAs, as well as how the tools will affect e.g. literary genres and the social organizing of literary life.

In developing (general) markup languages and their DTDs, markup specialists are now, in a sense, aiming for the disclosure of the inherent essence of texts and documents. In order to be able to adequately handle possible future document production and architectures, it is of utmost importance in LIS education and research to closely follow these attempts and perhaps also to be part of them.

Document architecture as an object of inquiry into information technology

The use of traditional IR systems tends to place modern technology as a barrier to products of another technology. Modern technology is treated as a mere tool for retrieving representations of information residing in a framework of print technology. This may be the most fruitful way of interacting with information systems in everyday practice. However, this apprehension may be of less use when learning modern information technology as such.

Mastering information technology, for example word processing, is not the same as to have an *understanding* of it. It is possible to learn word processing and in the process acquiring no or little understanding of how technology makes word processing possible. This results in problems in some cases, as in the exchange of documents between two subsequential versions of the same word processing software or when changing to another word processing software. The transition from Wordperfect to MS Word has for most users not been an easy one, even though there are apparent underlying similarities between their respective architectures. It is an interesting question to ask, *why* mastering the techniques of particular software and hardware solutions isn't enough and *if* it really isn't. We do not believe that it is, and the answer may be approached from different perspectives.

One of the reasons is the fact that modern information and communication technology conceals a course of events that takes place in the performance of certain tasks. Hidden under sophisticated and supposedly user-friendly interfaces are rather primitive but numerous events. It would be possible to say that this essence in modern technology tends to alienate its users, as was a frequent argument during the 60s and 70s, when computers came to be symbols for "big brother" (13). Be that an exaggeration or not, in the context of education this nature of concealment is certainly problematic.

If Bruner's statement that "the heart of the education process consists of providing aids and dialogues for translating

experience into more powerful systems of notation and ordering" (Bruner, 1966, p. 21) is worth considering, and we believe it is, then a major task will be to unveil the concealed courses of events in a way that facilitates the experience of technological foundations. The reason for the student's difficulties to understand concepts such as *file*, *server* or *network* may be that they, essentially, are elaborated abstractions. Consequently all those terms turn out as abstractions of unexperienced phenomena. Then the teacher finds him- or herself in a position where s/he has to tell beautiful stories the student hopefully will believe and remember, and/or instruct the student on how to perform a discrete and purely instrumental task.

There is no true magic inherent in technology. It's just that all hardware and software are built on a great amount of concealed abstractions. Several of these abstractions are founded on different theories on the nature of language and knowledge, as many investigations show.(14) Is it possible that if the underlying assumptions of for example representation techniques could be experienced and studied in a situated task, a thorough understanding would become possible?

Word processing is a common task that every student is expected to carry out. Problems with student's motivations and lack of interest in learning word processing, so often observed in other courses, are often absent. There is simply an obvious relation to necessary tasks. Word processing as a means to technology education should therefore seem like a good starting point for understanding technology. However, it is a fact that problems encountered during these tasks are proprietary to particular products. Even though several tasks in for example MS Word will be performed similarly to those in other word processors, they are seldom identical. The task of text alignment or line break in general can not be learned by the use of one word processing application alone. Another example is the insertion of in-line images, which can be done in several ways, described by software designers in ambiguous dialog boxes of the interface. How is the first-time user expected to recognize the difference and its consequential meanings between *storing* the image in the document as opposed to *linking* to an image, when both ways seem to produce the same effect? Technology and the solution of tasks tend to stay intimately intertwined, which makes it almost impossible to decontextualize problem solving and make way for learning, or with the words of Bruner again, technology hinders "independence of response from the immediate nature of the stimulus". (Bruner, 1966, p. 5)

A study of explicit DAs in relation to a task like this one would possibly be a remedy to this restrained growth. As mentioned in the beginning of this article, DA deals with "rules for the formulation of text processing application", but rules are not amenable nor explicit in the context of word processing applications. Such rules are in fact found among the general markup languages (GML) like SGML, XML and HTML, where technology may be stripped bare of its sophisticated interfaces.

GML reveals a lot more of what is going on "inside". Users frequently delight in seeing that a simple string of short text may cause an in-line image to appear in the text, or to see that another string ties an object to a different web site by a blue underlining. The same applies to changes made in a separate style sheet that may affect the colour and font of all first headings in a large complex web site. This is like *playing* with technology, not just using it as *a means to an end*.(15)

The need for DA implementation

There are numerous practices within LIS education and research where DA studies might prove to be an essential tool for gaining knowledge. A few instances have been addressed in this article, but our scope might of course be broadened to even further areas of LIS, for instance: a) the ever growing need for quality aspects and critical analyses of web contents and forms, and b) the overwhelming, new situation in digital environments, where the malleable screen presentation form of an object is separated from its storage form, in which two cases an understanding of the fundamental DA is crucial. But the study of DA should not be conceived of as a sub-discipline of electronic textuality solely. Extending this study to objects of earlier technologies may be revealing for the history of writing as a whole. The introduction of the title page may for example reflect important historical circumstances or changes in the technologies and the infrastructures of that time. Moreover, a thorough understanding of how different elements in different document types play particular roles, may serve in the work of cataloguing.

A large number of traditional institutions and systems in library environments are encountering obstacles in the digital world on account of their being essentially print-based. The established systems can be "bent" to

accommodate the qualities of digital documents - or the other way around. The obstacles can thus temporarily be by-passed through ad hoc-arrangements, but will not be adequately handled, unless solutions and theories are generated on the basis of thorough understanding and analysis of the architectural essence of documents in general and, in the examples stated, of web environments and digital objects in particular. This makes way for an architectural understanding of *different* kinds of media, structures and meta-objects. LIS education and research ought therefore implement the field of DA in order to perform necessary investigations of existing and future developments of documents and meta-objects. Students would be better prepared. LIS itself would be better prepared.

Notes

(1) The term *meta-object* is borrowed from William Arms (1997) and his colleagues' description of an "architecture for information in digital libraries". The term is for our part meant to indicate that many objects that serve as pointers to library material are not really bibliographic.

(2) The term *media chauvinism* has been borrowed from Espen Aarseth (1997).

(3)The reassessment needed pertains, of course, largely to the challenged concept of *document*, a concept "rooted in hundreds of years of tradition, planted firmly in enormous and complex systems for publishing, organization, and access. Yet it has become increasingly evident that the archetypal concept of 'document' as 'book' underlying these systems is insufficient to deal with a multitude of media formats, particularly diverse electronic formats (...)". (Schamber, 1996, p. 669). A retrospective survey of the difficulties involved in defining the concept of document is presented by Buckland (1997).

(4) It may be important to point out that even though the SGML standard dates back to 1986, work on markup languages based on separation of content and presentation dates back to the late 1960s, according to one of SGML's developer's, Charles Goldfarb (1998).

(5) Daniel Chandler (1995) attempts a thorough description and classification of writing strategies, where his "architectural strategy" is most akin to that implied by general markup languages.

(6) What has been said about meta-objects may also apply to this now widely used, at times even misused , term. Meta-objects *contain* metadata, but *are* not always bibliographic data.

(7) An attribution possible in using several search engines, e.g. *Alta Vista* and *Infoseek*.

(8) Of course, metaphors such as these are necessary elements of the architecture, as - along with navigation systems, indexing etc - "the glue that holds together a web site". (Rosenfeld & Morville, 1998, s. 11)

(9) That is: the attribution to the linguistic elements of a text with structural (and perhaps also presentationally intended) metainformation contained in "tags".

(10) A notable example of such a work within the humanistic disciplines is the TEI, *Text Encoding Initiative*, resulting in a particularly elaborate DTD (as well as a more manageable "light" version, the TEI Lite) (Sperberg-McQueen & Burnard, 1994).

(11) A task that might prove to be difficult indeed. When digitizing paper-bound works and their textual manifestations, we frequently have difficulty identifying, explaining, and explicitly tagging a great number of features in such works, as we have grown so accustomed to them that they have become more or less transparent.

(12) For instance printed drama, see Huitfeldt (1999).

(13) At another end of this chain is the practice of commercial software producers to "protect" from the user some patented layers of manipulation, thereby, in a sense, rendering the user impotent. This practice has been analyzed and criticized by e.g. literary theorists such as Friedrich Kittler (1994).

(14) See for example Richard Coyne's (1995) thorough exposé on the theoretical foundations for computational design or the works of Joseph Weizenbaum (1984), Terry Winograd and Fernando Flores (Winograd & Flores,

1986).

(15) Which, peculiarly enough is something that's not allowed in some of our public libraries, where actions taken by the help of computers must be good for something else or, as they put it, a task of searching.

References

(All web references checked March 31, 1999.)

* Aarseth, E. (1997) *Cybertext : Perspectives on Ergodic Literature*. Baltimore: Johns Hopkins UP.
* Arms, W. Y., Blanchi, C. & Overly, E. A. (1997) "An Architecture for Information in Digital Libraries". *D-Lib Magazine*, **3** : 2.
* Brown, J. Seely, & Duguid, P. (1996) "The Social Life of Documents". *First Monday*, **1** : 1.
* Bruner, J. (1966) *Toward a Theory of Instruction*. Cambridge, Mass.: Harvard UP.
* Buckland, M. (1997) "What is a 'Document'?". *Journal of the American Society for Information Science*, **48**: 804-809.
* Canfora, L. (1990) *The vanished library*. Berkeley: Univ. of Calif. Press.
* *CC:DA TF on Metadata and the Cataloging Rules: Final Report*. (1998) ALA, August 21, 1998.
* *The Center for Retrospective Digitization, Göttingen State and University Library* (1999) Göttinger DigitalisierungsZentrum, Niedersächsische Staats- und Universitätsbibliothek Göttingen.
* Chandler, D. (1995) *The Act of Writing: A Media Theory Approach*. Aberystwyth: Prifysgol Cymru (University of Wales).
* Coyne, R. (1995) *Designing Information Technology in the Postmodern Age*. Cambridge, Mass.: MIT Press.
* DeRose, S., Durand, D., Mylonas, E. & Renear, A. (1990) "What Is Text, Really?". *Journal of Computing in Higher Education*, **1** : 2 (winter), 3-26.
* Goldfarb, C. (1998) *The XML Handbook*. New York: Prentice Hall.
* Gudivada, V. et al. (1997) "Information Retrieval On The World Wide Web". *IEEE Internet Computing*, **1** : 5, 58-68.
* Heidegger, M. (1977) *The Question Concerning Technology and other essays*. New York: Harper & Row.
* Huitfeldt, C. (1999) "Tekstkoding og tekststrukturer", in: *Datahåndbok for humanister,* edited by E. Aarseth. Oslo: Ad Notam, Gyldendal, pp. 123-146.
* Kittler, F. (1994) "Protected Mode", in: F. Kittler, N. Bolz and Ch. Tholen, *Computer als Medium*. München: Fink, pp. 209-220.
* Kuhlthau, Carol Collier (1993), *Seeking Meaning : A Process Approach to Library and Information Services.* Norwood, NJ: Ablex.
* Lindquist, M. (1999) "Not Your Father's References : Citations in the Digital Space". *Journal of Electronic Publishing*, **4** : 3.
* McGann, Jerome (1991), *The Textual Condition.* Princeton, N.J. : Princeton Univ. Press.
* Pollock, A. & Hockley, A. (1997) "What's Wrong with Internet Searching". *D-Lib Magazine*, **3**: 3.
* Renear, A. (1997) "Three (Meta)Theories of Textuality", in: *Electronic Text : Investigations in Method and Theory,* edited by K. Sutherland. Oxford: Clarendon, pp. 107-126.
* Rosenfeld, L. & Morville, P. (1998) *Information Architecture for the World Wide Web*. Sebastopol, CA: O'Reilly.
* Schamber, L. (1996) "What Is a Document? Rethinking the Concept in Uneasy Times". *Journal of the American Society for Information Science*, **47** : 669-671.
* Simone, R. (1995) "The Body of the Text", in: *The Future of the book,* edited by G. Nunberg. Berkeley: UCLA Press, pp. 239-251.
* Sperberg-McQueen, C.M. (1991) "Text in the Electronic Age : Textual Study and Text Encoding, with Examples from Medieval Texts". *Literary and Linguistic Computing*, **6** : 1, 34-46.
* Sperberg-McQueen, C.M. & Burnard, L. (eds.) (1994) *Guidelines for Electronic Text Encoding and Interchange (TEI P3)*. Oxford & Chicago, TEI.
* Turner, R., Douglass, T., & Turner, A. (1996) *Readme.1st : SGML for Writers and Editors*. New York: Prentice Hall.
* Weizenbaum, J. (1984) *Computer Power and Human Reason*. Harmondsworth: Penguin.
* Winograd, T. & Flores, F. (1986) *Understanding computers and cognition*. Norwood NJ: Ablex.