Hypertext and Indexing Languages: 
Common Challenges and Perspectives

Abstract: Conceptual considerations on possible synergies between hypertext and indexing languages' research. The existence of a common conceptual and theoretical background encourages and facilitates cooperation. Researchers and developers in both fields face also common challenges in an environment of rapid technological and sociological change. The possibilities of crossed fertilization in both directions are carefully explored. It is concluded that the conjunction of the almost biological flexibility of the hypertext and the logical and conceptual consistency of indexing languages is a promise of better and more effective information management systems in the future.

1. Introduction

As it is well known, hypertext is an alternative to alphabetical and sequential access to information. Hypertext is also able to integrate dynamically quite different kinds of data—text, numbers, graphics, films and sounds—in what is called hypermedia. Finally, hypermedia seems to be the near future for all global information systems, due to the ever growing size and popularity of the WWW. But hypertext design may improve if the knowledge from classification and indexing research is applied. Inversely, old problems in documentary control may be solved with the aid of hypertext technologies.

In fact, both documentary languages and hypertext share a common problem—optimizing the representation and organization of universal and/or specific domain knowledge—and the same theoretical background, based on cognitive sciences and social epistemology. In both cases, their epistemological models rest on the constructs of concept and term, and the concept of relationship, that are called in hypertext systems nodes and links. Both of them rely on the same principle of non sequential access to information. Both of them are in pursuit of a universal organization of knowledge, like the universal classifications or Vannevar Bush’s Memex. Finally, both approaches have in common a flavor of utopia.

So, it is not strange that a basis for cooperation may be found in both directions. First, documentary languages—as vocabulary and conceptual control tools, and repositories of relational information—may provide the cognitive maps that can help us to fight two well-known problems of hypertext systems: cognitive overload and disorientation. Use of thesauri will lead to easier and wider retrieval and to improved automatic generation of hypertext relations. Classifications, on the other hand, will provide the universal frame for true conceptual retrieval, not simply an alphabetical one. This is of the greatest importance in a biologically growing environment like the Web, whose growth is not lead by central planning.

Inversely, hypertext opens new possibilities to the different documentary languages. First, hypertext can improve the interfaces on which the automation of indexing and classification systems relies. Second, the two-layers' theory for hypertext systems can provide a frame for considering the connection of the rational and pre-fixed lexical structures of the documentary languages and the idiosyncrasy and personal cognitive maps of the users. Third, hypertext can be
used to link associative and hierarchical languages, and those with texts and references, creating truly global and integrated information systems. Also, hypertext interfaces can facilitate the tools for designing and maintaining thesauri and classifications.

Finally, indexing theory and hypertext offer not only a potential for cooperation, but also share common challenges. Both of them require increasingly dynamic systems for logical and graphical representation, able to cope with the multidimensionality of the worlds of concepts. Both of them have as a common goal the development of platforms for cooperative and concurrent work. Both of them need to integrate automatic knowledge and linguistic processing. And finally, both hypertext and indexing theory are to gain a deeper understanding of the complementarity of the hierarchical and associative approaches to the organization of information systems.

2. A Common Problem and Theoretical Backgrounds

a) Though, as we have actually said, documentary languages and hypertext systems seem at a first glance two completely different things, both of them share a common underlying problem: optimizing knowledge representation, organization and transfer as a response to the information overload that our societies suffer.

In this sense, they are nothing more than two knowledge technologies, that find a common ground in the theoretical disciplines that study their very same problem, this is, cognitive sciences, Epistemology, etc.

b) The response to this common problem has been in both cases the development—more precisely, its importing from other sciences devoted to basic research—of a common theoretical background on the way that human beings acquire, represent and store information in the form of knowledge (from Cognitive Psychology) and how the society manages, co-ordinates and determines this knowledge (from Social Epistemology).

In fact, information scientists have given attention over all to one field of Cognitive Psychology: the study of the writing and reading processes. But information scientists have also made an important contribution to Social Epistemology: they have created an image of the Society of Knowledge as a macro-text, linked like a web by the conceptual relations existing among terms and the very genealogical relations existing among the documents themselves (The society of text, 1989).

c) In both cases, their core epistemological models are based on the constructs of concept and relation, which are called in the hypertext terminology nodes and links. As in thesauri, hypertext relations may be single or bi-directional, and also hierarchical or associative.

These relations can be labeled in quite more complex ways, as occurs in some quite complex documentary languages, like PRECIS.

Generally speaking, hypertext is an effort to transcend the limits of sequential communication and alphabetically organized information systems. Documentary languages achieve also this goal by means of their a priori structure.

d) Certainly, documentary languages and hypertext share a common core principle: non-sequentially in information access.

That is what Bush (1945) called "associative indexing" and what he proposed as the main feature of his "Memex", that is, that each item may be linked to any other, and the other way, too. Both systems aim at providing the universe of knowledge, which is mainly in a textual and sequential form, with a meta-textual organization, able to integrate every text in a super-structure that would surround, surpass and transcend it.

e) The above-mentioned reasons explain also that both disciplines use a common language.
As Churcher suggested in 1989, hypertext specialists have adopted the very language that it is used in Psychology, Artificial Intelligence, Pedagogy or Information Science—the so called Cognitive Sciences—to describe knowledge and information structures.

f) Finally and differently from other information technologies, both systems share a strong utopian component.

Certainly, hypertext systems, on one hand, and universal classifications and inter-connecting indexing languages—as the BSO, for example—on the other, aim at building tools for global knowledge organization. Not other was the driving idea of Dewey, Ranganathan, Bliss, Bush or Nelson.

Anyway, indexing languages and hypertext do not only share a common theoretical foundation, that leads them to similar views of the world of knowledge, but also, as we will see now, many possibilities of synergy and common challenges.

This conceptual and empirical proximity is not strange if we consider some common historical backgrounds. For example, thesauri and hypertext systems have a precedent in encyclopedias, dictionaries and complex and multi-referenced texts.

3. Contributions of Documentary Linguistics to Hypertext

Hypertext creates well-known problems either to the administrator of the system, either to the final user. These problems are the cognitive overload and disorientation (Conklin, 1987). Documentary languages are, precisely, the tools that can help us to overcome this handicaps. Indexing languages are tools for terminological control and they are also deposits of lexical relational information. Because of these, they conform true knowledge maps and models for conceptual organization that may improve navigability in hypertext environments.

The possible contributions of indexing languages for better hypertext systems can be summarized in these four points:

a) Terminological (or vocabulary) control. According to Frank Halasz (1988), hypertext systems should be able not only to provide with a navigational access to information, but also with suitable search and retrieval tools. Indexing languages, specially thesauri, are the tool for controlling search terms. In general, hypertext systems may get profit from every form of vocabulary control, from the pure morphological to the lexical ones. Thus, for example, Hyperlexis (Otal et al., 1992), which is a prototype of an interactive hypertext dictionary for lexical learning, has three kinds of retrieval tools: keyword search, a database of lexical relations for each word, and an assisted hypertext with a parser that leads from the clicked word—whichever its morphological form may be—towards its entry.

b) Automated link generation. From the point of view of vocabulary control, documentary languages can constitute an excellent knowledge database for automatic node identification and link building. Considering the costs of manual link creation, using databases of technical vocabularies with relational information is the only alternative to build massive hypertext corpora with minimal conditions of profitability (Westland, 1991).

c) Knowledge organization (structuring of information). As soon as 1991, Urr prevented us against the limitations of hypertext systems. They lack a coherent and—as Fugmann (1994) would call it—predictable underlying conceptual system, which is necessary to achieve one of its more important aims: giving quick and effective access to information with independence of the intellectual background of the user.

In his paper, Urr considered an analogy between hypertext and OPACs, which are less effective when they lack carefully developed classification schemes. Nowadays, an almost general agreement exists that a well-build hypertext system must have a structure able to lead users...
without needing any kind of formal or informal teaching about its use. In pursuit of this aim, several authors have proposed developing hypertext models with two or more levels, each of them teaching and giving access to the next and more difficult one (Agosti et al., 1992; Salminen and Watters, 1992). Documentary classifications could serve as conceptual superstructures, being common to the different local systems, and configuring a common cognitive interface to the different hypermedia products. In fact, multimedia products have been somehow criticized for their excessive idiosyncrasy and the low predictability of their relations. In this sense, this approach could complement other hierarchical approaches based on statistical models, as the one proposed by Botafogo et al. (1992).

d) Finally, as world-wide virtual hypertext systems develop in the Internet, completely new challenges in the field of documentary control are appearing. A brilliant example of this fact—nets of hypertext documents distributed throughout different computers along the globe—is the World Wide Web. As these nets grow, they will need progressively stronger knowledge organization tools and, of course, a tough work of terminological normalization. This is a new field for Knowledge Organization experts, and a very promising one, in fact. The possibility of studying logs of HTTP connections, for example, opens completely new fields in the empirical validation of knowledge organization structures.

Though all these are obvious considerations for any information scientists, very little work has been made in the above mentioned directions. As Gibbs (1992, 40) points out: "Hypertext is a technology which has caught the imagination of information scientists and authors alike. However, as with the first phases of expert systems, it has probably been oversold as a solution to all information production and retrieval problems. It has been shown to be effective in assisting users who have ill-defined information needs, or wish to discover information, but is generally less successful for goal-oriented searches. This is a direct result of the fact that hypertext research has concentrated on the preservation and creation of the explicit intellectual and structural links between text units which facilitate browsing. Despite observations concerning the potential of document surrogates for creating hypertext links (Reimer and Hahn, 1988), the indexing of text units has been generally ignored or conveniently overlooked by hypertext developers."

4. Contributions of Hypertext to Classical Knowledge Organization

Reciprocally, the concept of hypertext is offering new possibilities of controlling the vocabulary relations that traditional documentary languages provide. In current practice, non expert users find documentary languages—both thesauri and classifications, and, of course, more complex systems—difficult to use. Novel users are hardly able to take advantage of their conceptual rigor and their possibilities for better information retrieval. We think that hypertext environments can constitute the perfect platforms for the development of tools for both the professionals—modules for automatic management of thesauri and classifications, and the users—friendly retrieval interfaces.

Thus, hypertext systems might contribute to knowledge organization tools with the following advantages:

a) Make easier accessing to and using pre-coordinated languages. Hypertext systems can be the base for the friendly user interfaces to the documentary languages, especially the pre-coordinated ones, because hypertext interfaces usually combine keyword searching, alphabetical browsing and powerful relational navigation. Thus, for example, Bjorklund (1990) has studied the advantages and problems of using a hypertext system to make easier the use of the Swedish SAB for both users and professionals.

It is necessary, anyway, to have in mind that hypertext is more than a mere friendly user
interface technology. It is an alternative and complementary model in the access to information, apart from that of sequential, alphabetical or conceptual browsing. It is necessary to make clear the distinction between live hypertext, capable of storing (authoring) new relations, and fossilized hypertext, the set of relations that are provided by the creator of the system. It would be a pity to reduce hypertext to a mere 'wysiwyg' interface.

b) Connect the indexing languages with the cognitive maps of the users. As we have said before, a traditional problem with documentary languages is the difficulty to link their pre-defined structures—featured by their excessive crystallization and dependence from the ideology and intellectual backgrounds of their creator—with the idiosyncrasy and personal conceptual maps of the user, that differ form one person to the other, depending on its previous intellectual background, professional group, etc.

In this sense, hypertext has something new to offer to Knowledge Organization: the concept of path, which is a bridge between socially accepted knowledge and the individual processes of knowledge acquisition, which is always constructed in a personal and idiosyncratic manner (Novak and Gowin, 1984). In this sense, hypertext, as a living documentary system, suggest the storing of knowledge and learning paths from different points of view (disciplinary slanted, project slanted and so on). This would open a new dimension in the empirical research on indexing tools: it would be possible to compare the different paths established by users among them, and all of them with the indexing language. Stotts and Furuta (1991) have conceptualized the dynamic relation that exists in hypertext between the objective conceptual maps proposed by the creator of the system and subjective ones as the liking of two planes: the fixed one developed by the designer of the system—which is called by these authors fixed underlying information structure—and the flexible one, generated on the march (dynamically) by the user as it interacts with the system. The relation between documentary languages and the cognitive maps and linguistic uses of the users could be conceptualized in a similar manner. Thus, the documentary language could be adapted to the preferences of an information center or a group of users and even to a special user, but preserving an underlying vocabulary control and a common conceptual and lexical standard.

c) Dynamic interconnection of documentary languages among them, and of these with primary documents and references. Hypertext can be very useful to connect dynamically documents and their representations in the documentary systems, and thereafter with the documentary languages that give access to them.

As Alberico and Micco (1990, 180) stated: "Hypermedia provides the ideal tool for mounting and displaying bibliographic tools such as encyclopedias, Dewey and LCSN subject headings, and reference tools that can serve as guides to other information sources". Of course, hypertext interfaces for OPACs already exist, for example HYPERCAT of the Linkoping University in Sweden (Hjerppe, 1986). We have to keep in mind, however, that hypertext interfaces have not done as well in situations where there is a large, unfamiliar, heterogeneously structured network, such as traditional library databases, in which the user easily gets lost in the hyper-space (Alberico and Micco, 1990, 182). Use of hypertext in these environments requires ensuring the user with ‘intelligent’ help, as it happens in RABBIT (Tou et al., 1982).

Finally, hypertext systems could probably be used for joining dynamically classifications and thesauri among them, creating little by little some sort of interconnection languages.

d) Design, building and management of thesauri and classifications. Hypertext environments can be used for developing and managing thesauri and classifications: its terms, relations and graphical and alphabetical presentations (Richard, 1990).
5. Common challenges

As we told in the introduction, indexing languages' experts and hypertext designers share some common challenges of great importance, too. Here are some of the most important ones:

a) Development of dynamical systems for logical and graphical representation of the relations into and among documents. From a merely documentary point of view, one of the best strategies to improve information retrieval with keywords and descriptors has been to include them within a context. This context can be a priori, like lexical relations in thesauri; a posteriori, pre-coordination of indexing terms; or a mixed solution. Creating context throughout classifications and conceptual maps is also the best tool against cognitive overload and disorientation of hypertext users.

From a more theoretical point of view, documentary languages and hypertext systems face the same challenges in the logical and graphical representation of very complex conceptual systems. Especially, if we consider that this representation must be dynamic, because it refers to a world that is changing continuously. The map of sciences and the relations between disciplines are changing all the time, and so subject fields and their terminologies.

The world of graphical representation of concept maps is a very interesting field where we have a lot to learn from mathematicians (Wille, 1984). One of the most central problems is to conciliate the complexity of the relations among concepts and easy-to-use and navigable graphical presentations. We have to do also more research in which are the most suitable graphical presentations from the point of view of the users. It seems that presenting the more immediate relations of the node that is being consulted is more useful for the users than the presentation of the whole classification schedules (Girill et al., 1991).

b) Compatibility and hospitality for concurrent and cooperative work. A typical critic that both hypertext and associative documentary languages have deserved is that they have a very idiosyncrasy character. Each hypertext system is different from the rest of them. And that occurs also among, for example, thesauri. This results in problems when retrieving and interchanging information and while doing cooperative work. This is a very important question with indexing languages, because their magnitude needs of the cooperation among great groups of experts, and a fitted balance between the centralization of control functions and the respect of the principle of subsidiarity. These needs could be better served by the possibilities offered by cooperative electronic edition and distributed databases under a common interface.

c) Incorporation of automatic knowledge processing. The incorporation of inference engines to these structures of objects that are, after all, both indexing languages and hypertext systems would be able to transform them in dynamic retrieval systems. They would be able to assist the user in the search process and the professional in the indexing and classification processes, specially when determining complex relations.

In this sense, expert systems, on one hand, and indexing tools and hypertext, on the other, can complement each other: Expert systems shells provide the technology able to support inference processes, while indexing and hypertext systems store conceptual and pragmatical knowledge.

d) Deeper research in the complementary between hierarchical and associative approaches to knowledge organization. A typical aspect of the evolution of knowledge organization research has been fluctuation between associative and hierarchical models. But nowadays there is a broader sense of their complementary. In fact, the first conceptualizations of hypertext where based in an extreme form of the associationist theory of human thinking and memory. This was the case of Bush's Memex.

Recently, most of the authors highlight the relevance of a strong hierarchical organization of hypertext systems. Inversely, documentary languages have evolved from extremely rigid
tree-structures—like the first universal classifications—toward the assumption of associativity and multi-dimensionality of the world of knowledge in the thesauri and facet classifications. Certainly, from a broader point of view, it can be postulated that classification and association are complementary cognitive strategies in the task of representing and processing into knowledge (García and Esteban, 1993).

In this line of complementary between associative and hierarchical approaches to knowledge organization, it is not strange that Jan Wyllie (1990) defended recently that the future of hypermedia relies on the integration of free-text retrieval technology and of, in a broad sense, the know-how and theoretical basis of the discipline of documentary classification.

6. Conclusion

The conjunction of the almost biological flexibility of the hypertext project and the logical and conceptual consistency of indexing tools is a promise of better and more effective information management systems in the future. This is why working for a closer cooperation among researchers in both fields is so needed. While this occurs, research is needed in two directions. On one hand, we need to improve the theoretical background of both information technologies and connect their common references and referents with basic research carried on in the field of Cognitive Sciences. On the other, we have to implement practical models of cooperation between hypertext and knowledge organization tools.

Notes

1. Aims of such a system are quite meaningful: allowing the browsing of records and the navigation among them; building better structured catalogs than the traditional ones; offering alternative systems to show and illustrate the record structures and their information and relations; developing tools for establishing links and navigate among them; creating a dynamic system in itself, not only by continuously adding new records; ensuring each user the possibility of building and storing its own navigational routes throughout the library, so that the system itself can be enriched by the utilization that users do of it; allowing the storage of user models, both stereotyped and individualized; enriching the traditional information that catalogues offer, through relations among fields, records and files; and finally, offering information not only about individual documents, but also about the collections themselves.

References


