Innovation and Tradition in Knowledge Organization Schemes on the Internet, or, Finding One’s Way in the Virtual Library

Abstract: Because of the nature, number, and extreme mobility of resources available on electronic networks, access instruments that will allow information seekers to find relevant Web pages or sites are essential. While virtual libraries without clear collection development policies continue to be established, ad hoc knowledge organization schemes also continue to be conceived and implemented; these so-called innovative classification tools are generally structured around loose categories representing a mix of disciplines, subjects, forms of presentation, and audiences. This paper discusses a few examples of Web-based organization schemes, suggesting that their degree of intuitiveness, user-friendliness and efficiency is very possibly overrated.

1. Introduction and Background

In stark contrast with the rapidly fading image of the information superhighway, that of the “stuff swamp” (Crawford, 1999) is being proposed as an alternate representation of the Web environment. On the Web, finding a known resource whose exact address is not at hand, or finding relevant resources whose existence is not yet known, is a complex endeavour involving many pointless trips to a surprisingly large number of irrelevant locales. Extraordinary efforts are being made to bring some order to the chaos. The Web has been opened to increasingly wide segments of the general population, and this has led to the proliferation of attempts at providing thematic access to the resources that it contains. A majority of Web specialists maintains that it is not possible to organize the Web in the same way that libraries have been organized for centuries, because of the nature, number, flexibility and extreme mobility of network resources. Yet, these same characteristics render indispensable the implementation of logical and efficient instruments and strategies if we are to eventually find what we are looking for on the Internet. Weinberg (1996; 1999) and Crawford (1999) are among information specialists who do not believe that there is such a great difference between traditional documents and collections, and virtual documents and collections. Both rightly deplore that subject metadata is paid so little attention during discussions relating to Web resources.

A minimal amount of “physical” and conceptual organization of Web resources is now deemed essential by many. Why? For the same reasons that precluded leaving unorganized a traditional collection when the number of resources had reached a threshold beyond which it became difficult and costly to use. Web organizers gradually come to recognize that more and more individuals who search for information on the Internet have little time, and little patience, to spare.

Successive generations of information workers have tried, with more or less success, to solve well-known problems relating to the organization of resources with a view to their retrieval. Over the past 150 years, universal classification schemes, lists of subject headings, and thesauri have been used to order the masses of documents rapidly accumulating in physical libraries and more recently in databases worldwide.

When it came to organizing the Web, however, traditional methods were not deemed appropriate and/or necessary. To traditional discipline-based document classification schemes were preferred category-based systems, the categories being disciplines or themes (e.g.,
Education, Leisure), forms of presentation (e.g., Reference works, Dictionaries), potential audiences (e.g., Children's materials, New readers), specific place names (e.g., California), etc. Structured inventories of standardized terms were ignored in favour of keyword searches in the full text of a resource or in the metadata. Decision-making arguments invoked economic, time-related, and expertise-related justifications. Furthermore, existing organizing tools were judged too complex and inflexible for general use in the 21st century.

This paper presents and critiques a few examples of thematic organization of resources in virtual libraries residing on the Web. With so little space available, our goal is not to demonstrate that such or such way of organizing is best, but rather to feed a reflection on the necessity for true analytic processing and sound organization of the virtual library, both essential for the retrieval of a relevant resource when a reasonable effort to find it has been made by the user who needs it but does not yet know of its existence.

2. The Virtual Library

In this paper, the virtual library is defined as an organized collection of selected active links to electronic resources of a general or specialized nature. The virtual library exists where three functions, also characteristic of traditional libraries, are fulfilled. These functions are: selection of appropriate and valuable resources; organization of the selected resources, and provision of access to the organized resources. The World Wide Web in its entirety is not considered a virtual library, as no selection takes place on the Web, and because all of the resources it offers are presented as being of equal value.

There are many differences between traditional and virtual libraries, the most important of which being the absence of personalized services to users of the latter who must find their way with no assistance through virtual collections dispersed around the globe.

3. Systematic Subject Access in the Virtual Library

There are three possible ways of providing thematic access to a library collection: the physical (direct) access; the systematic (classified) access; and the alphabetical (lexical) access. While traditional libraries provide all three types of access, the virtual library does not provide direct access to its collection, forcing the information searcher through a first level of more or less systematic conceptual organization before allowing her/him to view a particular resource. In this paper, we look only at systematic accesses to virtual collections. The following examples show that the systematic organization may be quite complex, and that it is likely to differ greatly from one virtual library to another.

Example 1: Département d'informatique du Collège Bois-de-Boulogne in Montréal, Québec (triton.collegebdeb.qc.ca/biblio.htm)

Access to this small virtual library specializing in computer science is through a simple alphabetic presentation of basic form categories, with resources identified by specific terms.

Bibliographies [Informatique] [Internet]

Didacticiels [Active Server] [Active Server] [Java en français] [Java Script français] [Langage C] [Langage C] [Visual Basic] [Visual Basic] [Web Design]

FAQ

This is a simple and seemingly user-friendly way of organizing a small specialized collection. It is unfortunate though that, here as in many other virtual libraries, it has been assumed that there was no need to provide any descriptive information on a resource since its full content
was available at the click of a mouse. At Bois-de-Boulogne, it is thus a complete surprise for the user—and not always a pleasant one—to discover what is at the other end of a link. The link Bibliographies [Informatique], for example, provides the results of an AltaVista search of both terms, i.e., well over 30 000 Web pages.

Example 2: World Wide Web Virtual Library, Linguistics (www.emich.edu/~linguist/)

The Linguistics section of the WWW Virtual Library offers an unremarkable array of sections, presented in no meaningful order: The Profession, Interacting with LINGUIST, Research and research support, Language resources, Pedagogy, etc. Under each section title, further subdivisions represent categories of resources.

*Research and research support*
- Papers, Projects, Bibliographies, Topic-oriented sites, Texts

*Language Resources*
- Languages & language families, Dictionaries, Regional information

Although this organizing scheme may be quite useful, many headings leave much to be desired in terms of clarity: Topic-oriented sites? Regional information?

Example 3. LexUM, Centre de recherche en droit public de l’Université de Montréal (http://www.lexum.umontreal.ca)

Some fields of knowledge and practice, Law for example, should be easier to organize virtually than many others as they are already fairly structured in the “real world”. Although the sub-categories and resource types will definitely be easier to name, it still seems difficult to organize them in a way that will show some logical pattern and facilitate access to resources users know should be there in a specific form.

*Law resources of Canada and Quebec*
- Case Law,
- Quebec Charter of Human Rights and Freedoms

*International law resources*
- International legal instruments

*Law Library of Canada and Quebec*
- Law of Canada and Québec,
- Specialized organizations
- International Law
- International legal associations

This excerpt was chosen to show a confusing way of distributing virtual resources. First, it is odd to see “International law” appearing under Law Library of Canada and Quebec when there exists a broad class International Law resources. It is also odd to find “International legal associations” separated from “International legal instruments” to which it certainly is related.

Example 4. Yahoo! (www.yahoo.com)

Web organizers have discovered the efficiency of a hierarchical organization format which leads users, progressively, towards specific relevant resources. They have adopted this type of organization with amazing enthusiasm.

Yahoo!’s categories naturally cover a much wider range of disciplines, themes and subjects than any of the previous sites. The librarians at Yahoo! divides the world of
knowledge in fourteen broad classes of unequal status (Education stands next to Entertainment) and variable types (Arts & Humanities, Reference, Regional).

An expansion to the second level in the organization of the discipline Education shows more of the same absence of distinction between various categories of resources, and lack of directions as to what aspect of a complex subject (e.g., a bibliography on bilingual adult basic education in Canada) has been given priority in the classification of the resources:

   Adult and Continuing education [296]
   Bibliographies [5]
   Bilingual [16]
   Distance learning [411]
   Journals [27]
   Standards and testing [62]

One does not yet know much about the nature and content of the resources to be found at the other end of a link, but at least knows from the start how many resources will be listed. Further exploration of any sub-category would show that resources are generally listed in no order that could suggest significance or value.

Example 5. Librarian’s Index to the Internet (lii.org)

The Librarian’s Index to the Internet (LII) does not appear either as a model of systematic organization, judging by the following excerpt from the listing of first-level categories:

   Arts
   Automobiles
   California [The only place name given at this first level of division]
   Computers
   Cultures (World)
   Current events [How long will an event stay current?]
   Disabilities
   Education

The Education category is further subdivided as follows:

   General resources
   Basic education
   Business schools
   Choosing a college
   Distance education
   Tests

It was “Distance learning” in Yahoo!, it is “Distance education” in LII; “Standards and testing” is now simply “Tests”; “Continuing education”, an essential branch in the field, is nowhere to be found. These variations are to be noted in the internal organization of a domain whose terminology has been standardized in specialized dictionaries and thesauri.

Example 6. Canadian Information by subject (www.nlc-bnc.ca/caninfo/esub.htm)

Not surprisingly, the National Library of Canada sticks to the true and tested when it comes to ordering its virtual library of Canadian resources. With a slightly refreshed and updated terminology, the Dewey Decimal Classification is brought back to serve in the new environment.
39

0 General subjects, general works
3 Social sciences
6 Applied sciences, mathematics

3 Social sciences
  30 Social sciences in general (incl. sociology, social groups) [343 links]
  32 Political sciences (including political parties, international relations) [177 links]
  37 Education [400 links]

37 Education
  370 Education
    Canadian Education Association
    Education in Alberta
    Infobourg de l'éducation

  370.113 Enseignement professionnel
    Inforoute FTP

  370.7 Éducation - recherche
    Canadian Society for the Study of Education

The number of resources available is given at the second level of expansion only. Titles of specific resources are seen at the third and last hierarchical level, attached to class numbers that are specific enough to considerably limit the number of links offered to the information seekers.

4. Discussion

Self-proclaimed organizers of Web resources in virtual libraries declare that the thematic organization they provide is more intuitive and user-friendly than that offered in traditional libraries. More likely, the effort to be provided by an individual in search of relevant information in a virtual library is not smaller, and may even be greater given the absence of personal assistance, than that which must be provided by this same user in the traditional library.

The flawless subject organization structure, never achieved in traditional classification schemes, is not about to be found in any of the innovative schemes that now abound on the Web, and it is difficult to perceive how the systematic organization proposed by Yahoo!, for example, would be more intuitive than that proposed by the Dewey Decimal Classification.

In the new schemes, categories are seldom mutually exclusive, thus generating ambiguity and confusion. Although any resource could and should be assigned to more than one category, the traditional practice of assigning a resource to a single class continues to dominate in the new environment, for strictly economic reasons.

The new knowledge organization schemes share important weaknesses:

- an obvious lack of concern for any kind of standardization, whether it be at the terminological or at the structural level;
- a lack of consistency in the arrangement of categories, subcategories, etc.;
- a total disregard for theoretical principles of knowledge organization (facet analysis, citation order, etc.);
- a lack of specificity, with users being led to long lists of unqualified and undifferentiated resources even at the lowest level of specificity allowed by the system, most likely a third level;
- a lack of true and sound relational structures, and of references from one subject to related ones.
5. Conclusion

What is the real reason explaining the under-use of traditional organization schemes in virtual libraries? We believe that ignorance may have played a major role in early rounds of decision-making. Classifiers at Yahoo!, for example, emphasize their use of a bottom-up approach to structuring hierarchies, opposing it to the top-down approaches supposedly favoured in traditional schemes (Callery, 1996). But in fact, none of the traditional document classification schemes currently in use still adheres to the strict top-down approach (the Library of Congress Classification never did!) It is a well-known fact that the best performing schemes are developed through a combination of approaches granting equal importance to literary warrant (what is offered) and scientific warrant (what is likely to be requested).

Sufficient time and effort has certainly not yet gone into the design of innovative organization schemes for use in virtual libraries. The category systems are already starting to crack at the seams and to experience many of the problems that afflict the older schemes (subjectivity, lack of currency, insufficient specificity, etc.) (Weinberg, 1996). It is becoming increasingly difficult to convince trained users that the new schemes are not the product of amateurish improvisation, that they are useful, and that they serve them well.

At this point in time then, innovative knowledge organization schemes still have to demonstrate that they are easier to use and more efficient than the old ones. But let's not forget that we also still have to demonstrate that the traditional schemes are up to the task of organizing the virtual library and are more efficient than the new ones, for reasons other than that of their long-standing and widespread use in physical libraries.

At this critical junction, the proponents of traditional classification schemes and the designers of category systems must join forces to study and get a much clearer idea of what the user of the virtual library needs, and of how this real user with real information needs would really like to access the collection.

References