Subject Organization in Three Types of Information Resources:
An Exploratory Study

Abstract
Knowledge tends to be structured differently in different types of information resources and information genres due to the different purposes of the resource/genre, and the characteristics of the media or format of the resource. This study investigates subject organization in three types of information resources: books (i.e. monographs), Web directories and information websites that provide information on particular subjects. Twelve subjects (topics) were selected in the areas of science, arts/humanities and social science, and two books, two Web directories and two information websites were sampled for each subject. The top two levels of the hierarchical subject organization in each resource were harvested and analyzed. Books have the highest proportion of general subject categories (e.g. history, theory and definition) and process categories (indicating step-by-step instructions). Information websites have the highest proportion of target user categories and genre-specific categories (e.g. about us and contact us), whereas Web directories have the highest proportion of specialty categories (i.e. sub-disciplines), industry-role categories (e.g. stores, schools and associations) and format categories (e.g. books, blogs and videos). Some disciplinary differences were also identified.

Introduction
In an information resource (e.g. a book or website), information is organized to serve the purpose of the resource, to make the content logically coherent and to help the reader/user to access and assimilate the information effectively. The overall information organization is usually hierarchical, and may be presented as a table of contents (in the case of a book), a menu system (for a Website), or a hierarchically organized set of hyperlinks (for a Web directory or the site map of a website). These give the reader/user an overview of the information content of the resource - especially the main topics and sub-topics and their relationships. They also support navigation and browsing, allowing the user to quickly zoom into a particular subtopic of interest.

Information on a particular subject is likely to be organized differently in different types of information resources and information genres to serve different purposes and different target users. The characteristics and constraints of the format or media may also impact the organization of information. The purpose of the study is to find out how information or knowledge tends to be organized in three types of information resources:

1. books on a particular subject, represented by their table of contents
2. information websites on a subject, represented by their menu systems
3. Web directories of resources on the Web, represented by their hierarchically organized hyperlinks to resources on a subject.

These resources have either a pedagogical purpose to impart knowledge about a particular subject, or a reference function to link users to a range of resources related to the subject. Thus information organization in these resources can be considered as knowledge organization of a particular subject or field, from a particular perspective. The knowledge
organization is assumed to be hierarchical and can be represented by a taxonomy (or tree structure).

A book (monograph) on a particular subject is limited to a few hundred pages and tends to have a pedagogical purpose to impart knowledge about a subject. Readers are expected to read the book from beginning to the end in sequence, or at least a substantial portion of the content (e.g. a chapter). A book is organized into chapters and chapters are subdivided into sections.

An information website provides information on a subject, and may serve a combination of pedagogical and reference functions. It is not limited by the number of pages and moreover has hyperlinks that link to other Web resources. Information is presented in smaller self-contained chunks that fit on a webpage that is two or three-screens in length. The reader is not expected to read all the pages or read them in a particular sequence.

A Web directory serves primarily a reference function. It does not have much native content stored on the resource, but has a large number of hyperlinks that refer the user to many different types of resources on the Web. It is likely to have a large knowledge structure because providing a knowledge structure is its main function.

In analyzing the information organization in the three types of resources, we attempted to identify recurring patterns and facets (category types) that occur across resources and differences between the types of resources. As the information organization may be affected by the subject area, we selected 12 subjects of different types - two broad subjects and two narrower subjects each from the sciences, arts/humanities and social sciences. Two books, two Web directories and two information websites were sampled for each subject. To keep the analysis manageable, only the top two levels of the knowledge structures (taxonomies) were analyzed. Kwasnik (1999) referred to this as the “first cut” that determines the overall shape and “representational eloquence of the classification.”

This is part of a series of studies we are carrying out on knowledge organization in different environments and for different purposes.

Previous Studies

There are surprisingly few studies of information/knowledge organization in various information genres and media. Most studies of websites have focused on usability and interface design. Hardly any have focused on the navigation structure, other than to determine the importance of ease of navigation (e.g., Parboteah, Valacich & Wells, 2009; Tarafdar & Zhang, 2005) and to present general guidelines for the design and presentation of the navigation structure (e.g., Oppenheim & Ward, 2006; Perugini & Ramakrishnan, 2006).

Chen, Magoulas and Dimakopoulos (2005) studied users’ cognitive styles and their preference for different kinds of hierarchical structures for browsing Web directories. They compared two cognitive styles: field independence and field dependence. Based on their study of three Web directories - Google, Alta Vista and Lycos - they concluded that field dependent users, who typically see the global picture and ignore details, prefer a wide and shallow hierarchical structure (with many main categories and few levels of subcategories), and prefer the main categories and subcategories to be presented on different screens. On the other hand, field independent users, who are less dependent on the context and tend to focus on details, prefer a narrow and deep hierarchy (with a few main categories and many levels of subcategories), and prefer the main categories and subcategories to be displayed on the same screen.

Kwasnik (2002) analyzed the classification scheme used in Amazon.com to organize books. She found a rich network of terms in an enumerative and multihierarchical structure” and multiple paths to the same books. At the top level, topical categories were
mixed with genre categories, and categories reflecting perspectives, formats and promotions/offers. There was no consistency in scale, granularity, terminology or structure. The structure was pragmatic - to increase the likelihood of a user making a purchase - rather than to provide a knowledge organization structure. Chang (2004) analyzed the types of category redundancies found in the classification schemes of Amazon.com and the U.S. Energy Information Administration website. She found that over half of the categories were redundant, and identified five types of redundancies that suggest different kinds of issues. The redundant categories support the pragmatic purposes of the websites by providing multiple entry points and paths to the same item.

In this study, the books, Web directories and information websites that were analyzed seek mainly to provide information, though some of the information is of services, organizations and personalities. We argue that the information organization structure in these resources reflect the author/designer’s knowledge structure, which includes a high-level model of the academic discipline (the first cut) as well as industry knowledge. The extent that this coincides with the user/reader’s conception probably varies. But for the information organization structure to be useful, it should be compatible with the user’s conception, or at least something the user can grasp. However, we concede that this issue needs further analysis and discussion, and we hope that the results of our study will contribute to this discussion.

Research Method

The selection of the subjects started with the list of academic disciplines in Wikipedia (http://en.wikipedia.org/wiki/List_of_academic_disciplines). Six broad subjects and six narrower subjects were selected, based on the authors’ interests:

1. From the Health/Medicine discipline:
   • broad subjects: health and medicine
   • narrower subjects: cancer and surgery
2. Music discipline:
   • broad subjects: string instruments and folk music
   • narrower subjects: violin and flamenco music
3. Education discipline:
   • broad subject: education
   • narrower subject: child education.
4. Psychology discipline:
   • broad subject: psychology
   • narrower subject: child psychology.

For each of the subjects, we selected two books, two Web directories and two information websites. Altogether, 72 (12x6) resources were analyzed. The resources were identified by carrying out title-keyword searches in a library catalog and searching for the subject keyword in Google.

From each of the resources, we harvested level 1 and level 2 category items from the knowledge structure of the resource: the table of contents for books, the hierarchical menu structure from information websites, and the hierarchy of hyperlinks from Web directories.

From a preliminary analysis of the knowledge structures, we identified seven recurring facets in the organization of information:

1. Specialty facet
2. General subject facet
3. Industry role facet
4. Process facet
5. Target user facet
6. Format facet

These are explained in the next section. They are used as a framework to analyze the similarities and differences in information organization in books, Websites and Web directories.

**Results**
Table 1 gives the distribution (proportion) of categories from the seven facets - for books, Web directories and information websites. Table 2 provides more detail, listing the average number of categories of each facet that can be expected in each type of resource. Of the three types of information resources, Web directories have the biggest breadth in knowledge structure with an average of 15.3 categories at level 1 and 31.0 at level 2 (see Table 2). Books are next with an average 12.9 chapters and 22.6 chapter sections. Information websites have an average of 9.0 categories at level 1 and 23.1 categories at level 2. The three types of information resources have different profiles in the distribution of categories from the seven facets. We discuss each facet in more detail.

**Table 1: Percentage of categories from the seven facets for the three types of resources**

<table>
<thead>
<tr>
<th>Facet</th>
<th>Book</th>
<th>Information website</th>
<th>Web directory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specialty facet</td>
<td>46.4%</td>
<td>40.7%</td>
<td>60.7%</td>
</tr>
<tr>
<td>General subject facet</td>
<td>37.1%</td>
<td>30.7%</td>
<td>14.4%</td>
</tr>
<tr>
<td>Industry role facet</td>
<td>2.5%</td>
<td>5.3%</td>
<td>12.2%</td>
</tr>
<tr>
<td>Process facet</td>
<td>2.9%</td>
<td>0.9%</td>
<td>0.7%</td>
</tr>
<tr>
<td>Target user facet</td>
<td>1.5%</td>
<td>8.1%</td>
<td>4.7%</td>
</tr>
<tr>
<td>Format facet</td>
<td>0</td>
<td>2.7%</td>
<td>3.9%</td>
</tr>
<tr>
<td>Genre-specific category</td>
<td>9.6%</td>
<td>11.5%</td>
<td>3.4%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

*The highest percentage in each row are highlighted in italics*

**Table 2: Average number of categories of each facet in an information resource, for the 3 types of resources**

<table>
<thead>
<tr>
<th>Facet</th>
<th>Specialty</th>
<th>General subject</th>
<th>Industry role</th>
<th>Process</th>
<th>Target user</th>
<th>Format</th>
<th>Genre-specific</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book Level 1</td>
<td>5.58</td>
<td>4.42</td>
<td>0.08</td>
<td>0.29</td>
<td>0.16</td>
<td>0</td>
<td>2.33</td>
<td>12.86</td>
</tr>
<tr>
<td>Level 2</td>
<td>10.87</td>
<td>8.75</td>
<td>0.79</td>
<td>0.75</td>
<td>0.37</td>
<td>0</td>
<td>1.08</td>
<td>22.61</td>
</tr>
<tr>
<td>Total</td>
<td>16.46</td>
<td>13.17</td>
<td>0.87</td>
<td>1.04</td>
<td>0.54</td>
<td>0</td>
<td>3.42</td>
<td>35.50</td>
</tr>
<tr>
<td>Information website Level 1</td>
<td>2.96</td>
<td>2.08</td>
<td>0.58</td>
<td>0.08</td>
<td>1.17</td>
<td>0.46</td>
<td>1.71</td>
<td>9.04</td>
</tr>
<tr>
<td>Level 2</td>
<td>10.13</td>
<td>7.79</td>
<td>1.13</td>
<td>0.21</td>
<td>1.46</td>
<td>0.42</td>
<td>2</td>
<td>23.14</td>
</tr>
<tr>
<td>Total</td>
<td>13.08</td>
<td>9.88</td>
<td>1.71</td>
<td>0.29</td>
<td>2.63</td>
<td>0.88</td>
<td>3.71</td>
<td>32.17</td>
</tr>
<tr>
<td>Web directory Level 1</td>
<td>10.54</td>
<td>2.25</td>
<td>1.08</td>
<td>0</td>
<td>0.46</td>
<td>0.50</td>
<td>0.42</td>
<td>15.25</td>
</tr>
<tr>
<td>Level 2</td>
<td>17.54</td>
<td>4.42</td>
<td>4.58</td>
<td>0.33</td>
<td>1.71</td>
<td>1.29</td>
<td>1.17</td>
<td>31.04</td>
</tr>
<tr>
<td>Total</td>
<td>28.08</td>
<td>6.67</td>
<td>5.67</td>
<td>0.33</td>
<td>2.17</td>
<td>1.79</td>
<td>1.58</td>
<td>46.29</td>
</tr>
</tbody>
</table>

*Note: The figure in each cell is an average from 24 resources*
**Specialty Facet**

The specialty facet refers to subdivision into sub-disciplines. Sub-disciplines are domain dependent—they are different from one subject to another. For example, medicine was subdivided into cancer, heart disease, surgery, etc. and music was subdivided into instruments, genres, etc.

Table 1 shows that specialty categories are the most common for all three resource types, accounting for 40% to 60% of the categories. Web directories have the highest proportion of categories by specialty, accounting for 60.1% of the categories in level 1 and 2. For example, the Web directory "Education Planet" has nine level 1 categories, eight of which are sub-disciplines such as math, art and music, and geography. This is probably because Web directories seek to be fairly comprehensive in its coverage of relevant resources on the Internet, and carries minimal native content on its site.

**General Subject Facet**

The general subject facet refers to common subject subdivisions such as history, definition and education that can be used to subdivide any subject. General subject is the next most frequent type of category after specialty, often found in books (37%) and websites (30%), and less frequently in Web directories.

These common subjects, e.g. history and definition, serve a pedagogical function and provide a foundation for understanding a subject. They are thus more likely to be found in books and information websites. For example, the book "Introduction to Education Studies" has chapters on "What is education studies?" (i.e. definition), "The nature of education" and "An modern history of schooling" at level 1. The book "Education Studies: An Introduction" has chapters on "Theoretical perspectives" and "Policy." On the other hand, Web directories serve primarily a reference function and have fewer categories devoted to introductory or tutorial material.

**Industry Role**

The industry role categories are practical in nature and refer the reader to available services. In the medicine domain, it includes hospitals, clinics and medical associations; the music domain has music stores, music schools and luthiers; and the education domain has kindergartens and internships. Industry role categories are mainly found in Web directories (12.2% of categories), and to a smaller extent in information websites (5.3%).

**Process Facet**

Process categories indicate content that provide step-by-step instructions. Examples in the medicine domain are "7 steps for cancer prevention" and "how to lose weight"; and the music domain has "learn to play violin," "how to make a violin" and "Suzuki violin method." Process categories occur most often in books (2.9% of categories) as the categories indicate a teaching function.

**Target User Facet**

Target user categories indicate information that is oriented to a particular target audience. For example, the medicine domain has categories for doctor" and "for patient"; the music domain has categories "for child" and "for adult"; and the education domain has categories "for teacher" and "for student." These categories occur most often in information website (8.2%), followed by Web directories (4.7%).

**Format Facet**

Format includes categories such as books, blogs, videos, albums and Flash. These categories are more often found in Web directories (3.9%) and websites (2.7%), but seldom in books.
Genre-specific categories

Genre-specific categories are categories that are associated with particular media or genre. For example, books often have preface, bibliography and index; information websites often have home, about us, contact us and site map; and Web directories often have links and personal pages. These genre-specific categories are more often found in books (9.6%) and websites (11.53%), and less often in Web directories (3.4%).

Differences Across Disciplines

We compared the proportions of the facets across the disciplines and found some interesting differences. The differences may be random effects due to the small sample size. However, they suggest issues to explore further.

We found the music discipline to have a higher proportion of specialty categories than expected and lower proportion of general subject categories:

- For music information websites: a high 65% of specialty categories (compared to 41% for all 4 disciplines) and a low 9% of general subject categories (compared to 31% for all 4 disciplines)
- For music books: a high 66% of specialty categories (compared to 46% overall) and a low 25% of general subject categories (compared to 37% overall).

In contrast, education and psychology have a lower proportion of specialty categories and higher proportion of general subject categories:

- For education/psychology information websites: a low 28% of specialty categories and a high 39% of general subject categories
- For education/psychology books: a low 39% of specialty categories and a high 51% of general subject categories.

This suggests that the music resources sampled are more specialized or address a more specialized audience, whereas the education and psychology resources address the general public. In fact, the education and psychology websites in our sample carry a substantial amount of educational resources for parents and teachers. This may reflect the nature of the disciplines or may be an artifact of the sampling.

The medicine and music Web directories in our sample contain a high 20% of industry role categories (with links to organizations, products and services), whereas the education and psychology Web directories contain a low 1% of industry role categories (with more links to educational resources).

Conclusion

This exploratory study has sampled only a small number of disciplines and a small number of resources. Though it is not possible to make statistical generalizations, the results are suggestive and do make sense. Books, Web directories and information websites have different information organization structures, according to the different purposes they serve. Books have a higher proportion of general subject categories and process categories as they serve a pedagogical purpose of imparting basic knowledge on a subject. They also have genre-specific categories such as preface, bibliography, appendix and index.

Information websites have a higher proportion of target user categories. As they are easily accessible on the Internet and are visited by users with different needs and levels of expertise, they may attempt to serve both experts (e.g. doctors) and lay users (e.g. patients). Websites also have a high proportion of web-specific categories such as about us, contact us and site map.

Web directories have the highest proportion of specialty categories (i.e. sub-disciplines) as they serve a reference function and attempt to provide an extensive coverage of relevant
resources on the Internet. They also serve practical functions with industry-role categories that help users to locate services, and have more format categories indicating coverage of resources of different media and formats.

These results need to be confirmed by a larger-scale systematic study. Disciplinary differences are likely to be found, and indeed some disciplinary differences were found in our sample.

The facets adopted in this study were derived rather informally from a preliminary review of the categories harvested from the sample resources. The facets appeared salient probably because of the researchers’ familiarity with the purposes and characteristics of the different information genres. The seven facets were found to be useful in distinguishing between the types of resources, but there are conceivably other ways of grouping categories into facets and other potentially useful facets. Focus groups can be carried out to identify other facets to investigate. In addition, frequency analysis can used to identify common categories which may reflect facets. Some facets are likely to be discipline dependent. Facets that are important in the music domain (e.g. genre, instrument and time period) may not be important in other domains. In this study, the categories were assigned to facets manually but no inter-coder reliability test was carried out. Making a list of categories under each facet will help to improve coding reliability.

As the specialty facet accounts for the highest proportion of categories, future work can examine this facet to identify differences in the way the different resources divide a subject into sub-disciplines. Some of the categories coded in this study as specialty categories can be considered to be separate facets, e.g. instrument and genre in the music discipline. It will also be revealing to compare the subject organization found in the different information genres with the knowledge structures in general classification schemes such as the Library of Congress Classification Scheme and Dewey Decimal Classification.

References