

THE COMPARISON OF HIERARCHICAL STRUCTURE OF SUBJECT DIRECTORIES YAHOO, THE OPEN DIRECTORY & LOOKSMART WITH DEWEY DECIMAL CLASSIFICATION IN 10 SELECTIVE SUBJECTS

Z. HAYATI, Ph.D.

Department of Librarianship and Information Science
Shiraz University, I. R. of Iran
email: zhayati@rose.shirazu.ac.ir

P. TAJER, M.S.

Islamic Azad University – Marvdasht Branch
Marvdasht, I. R. of Iran
email: pegtaj@yahoo.com

Abstract - This survey is aimed to compare the hierarchical structure of subject directories, including Yahoo, The Open Directory and Looksmart with Dewey Decimal Classification (DDC) in 10 selective subjects. Five questions are raised to compare the broadness and the depth of the hierarchical structures in the subject directories and DDC and to study the effectiveness of the usage of DDC in the organization of the Internet resources. The data were collected by counting links and indentions. The results show that the hierarchical structure of Looksmart is broader and deeper than Yahoo, The Open Directory and DDC. This characteristic decreases its effectiveness as a classification scheme. This study also shows that subdivisions are rather evenly distributed in different hierarchical levels of DDC in comparison to subject directories. These features in addition to their general nature make DDC a suitable instrument for organizing Internet resources.

Keywords - Dewey Decimal Classification (DDC), Internet Resources Organization, Subject Directories, Hierarchical Structure, Information Retrieval.

INTRODUCTION

The size of information on the Internet increases very rapidly. Although computer specialists try to organize the Internet resources, effective organization of information in this network is still a real problem. Using keywords for searching the Internet resources through search engines is a very popular approach. But many studies show that users often faced with unwanted and irrelevant items when using this approach [4].

Using classification or subject directories is another approach for searching among the Internet resources. In this approach, information is classified by subject and each directory displays several distinct subject categories. In fact, resources are set in several subject groups [2] which are very similar to numeric classification schemes such as Dewey Decimal Classification (DDC).

Librarians and information specialists have long recognized the potential of library classification schemes for improving subject access to information [5]. Therefore, when the Internet subject directories appeared as a retrieval tool, they offered the idea of using

library classification schemes to organize the Internet resources. This study sets its aim as to compare the hierarchical structure of subject directories such as Yahoo, The Open Directory and Looksmart with Dewey Decimal Classification. The comparison has been carried out in 10 selective subjects which have been derived from different schemes.

LITERATURE REVIEW

Several authors attempted to show the credibility of library classification schemes, like DDC in organizing electronic resources on the Internet. Balas [1] carried out a study to show the applicability of the DDC in organizing information on websites. He came across with several web sites using this scheme as an organizing tool. He stated that DDC is still dynamic and alive [1]. In order to explore the effectiveness of the DDC, Woodward [8] studied directories such as BUBL, Cyber Dewey and World Wide Web Virtual Library. He concluded that using library classification schemes can be useful if they have been improved to be applied to the Internet resources [8].

Wheatley [7] compared the hierarchical structure of six Internet subject trees which were Looksmart, Yahoo, Infoseek Go, Britannica, Excite and Lycos. He found out that subject trees based on library classification principles made browsing of information simple and easy when they were used as retrieval tools. He confessed that in the near future, this approach, in contrast to search engines, would become the only practical and effective way for organizing and retrieving electronic information [7].

Vizine-Goetz [6] went a little further and studied the hierarchical structure, the subject classification and postings in Yahoo, Looksmart and DDC. She stated that both subject trees and DDC provided users to have alphabetical and hierarchical access to electronic resources and DDC could effectively be used in big electronic collections [6].

RESEARCH QUESTIONS

In order to study the credibility of DDC as a classification scheme in the Internet environment we attempted to answer the following questions:

1. What are the hierarchical positions of the selected subjects in subject directories as compared with DDC?
2. How are the hierarchical structures in DDC and subject directories expanded?
3. How is the hierarchical depth in DDC and subject directories?
4. What is the frequency of subdivisions in the hierarchical levels?
5. Is it possible to use DDC for effective organization of the Internet resources?

METHODOLOGY

A survey study is carried out to answer the above questions. Hierarchical structures of

some parts of all DDC main classes are studied in this research. So one subject was picked up from each class of DDC and compared with symmetrical subjects in the directories. Ten subjects were selected as follows: Library & Information Science (020), Psychology (150), Islam (297), Economics (330), English Language (420), Physics (530), Agriculture (630), Paint (750), English Literature (820) and United States History (973). The data were collected by the counting of links and indentions. The numbers of subdivisions were also counted in different hierarchical levels.

DATA ANALYSIS

1. What are the hierarchical positions of the selective subjects in the directories as compared with DDC?

To determine the hierarchical positions of the selective subjects in the directories and DDC, tree displays were depicted. For example, the tree displays for Librarianship & Information Science show that level three in Yahoo and TOD includes this term (Figure 1). But level four in Looksmart and level two in DDC consist of this subject and its subdivisions.

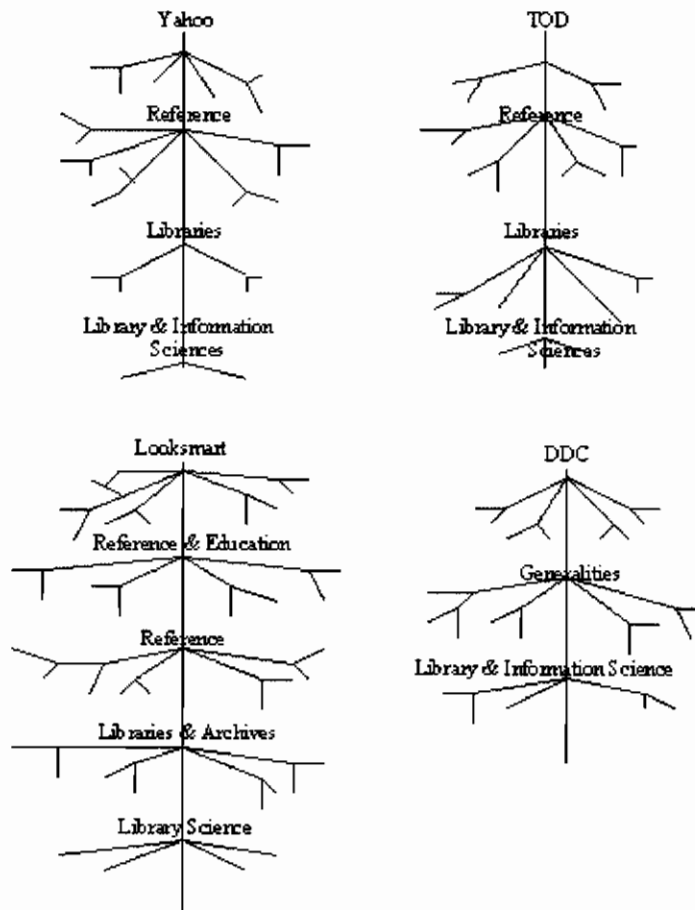


Figure 1: The tree displays of Librarianship & Information Science.

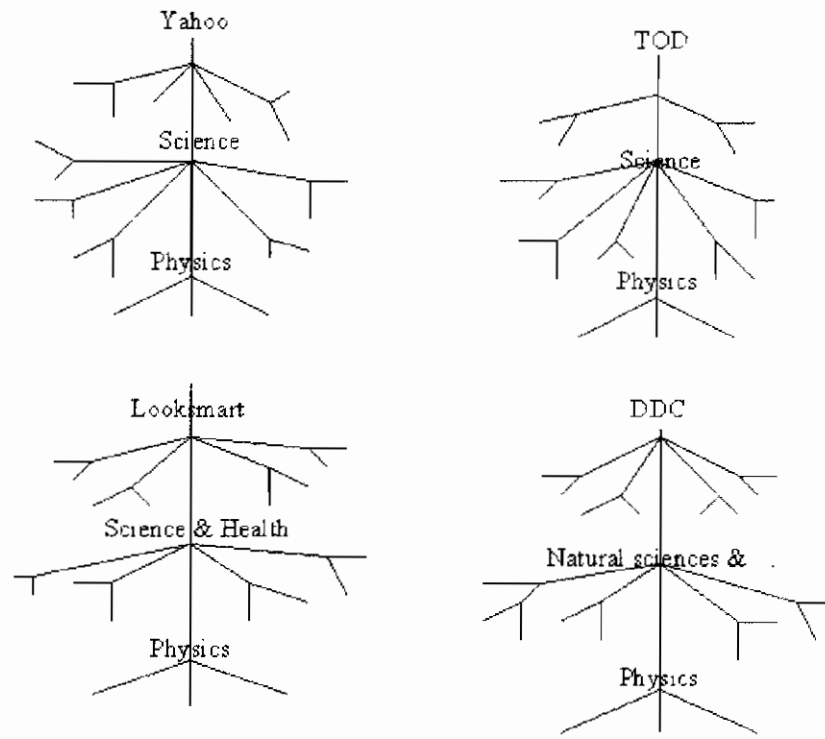


Figure 2: The tree displays of Physics.

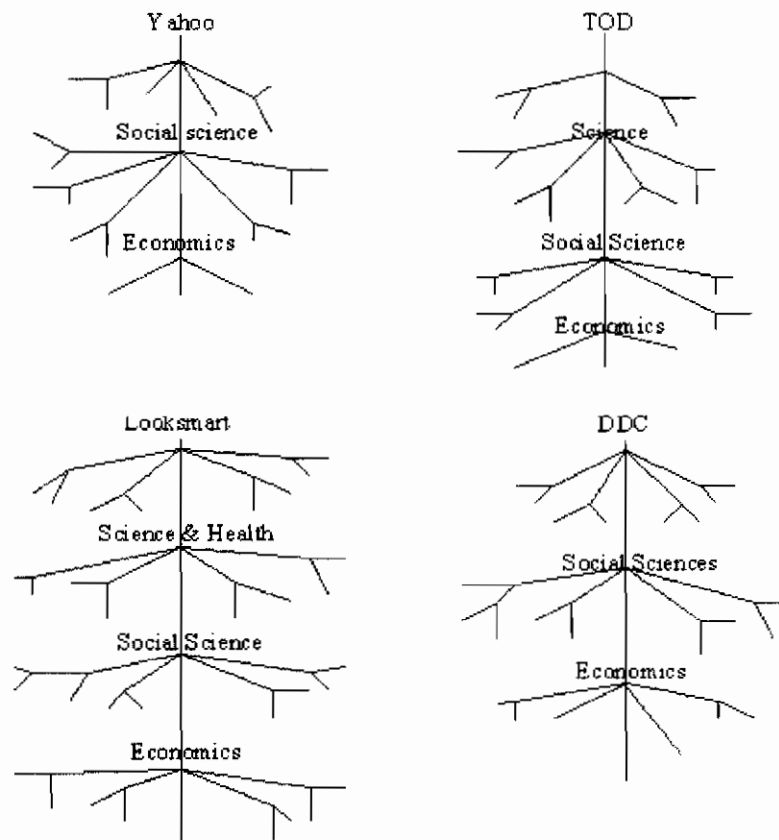


Figure 3: The tree displays of Economics.

The tree displays of the ten selective subjects show that although there are few similarities, the hierarchical structure of subjects in DDC and directories are not similar. The only subject that has the same position in DDC and directories is Physics, which is under the science category (Figure 2). Yahoo and DDC classified Economics as a subdivision of the social science category (Figure 3). The directories themselves are different in hierarchical structures. The only subject that has the same position in these three directories is Physics (Figure 2).

2. How are the hierarchical structures in DDC and subject directories expanded?

To study the expansion of hierarchical structures in DDC and subject directories, links and indentions were counted. Table 1 shows that Looksmart has the maximum total (1770.1) and average (177.01) subdivisions in the hierarchical structure. Figure 4 also shows that subject directories have more subdivisions in their hierarchical structures than DDC and this is because of the expansion of subject directories through the name of countries, people, plants, etc.

Table 1: The total and average subdivisions in the hierarchical structures of DDC and directories.

Categories / Directories	Yahoo	The Open Directory	Looksmart	DDC
Library & Information Science	21	132	176	318
Psychology	123	230	1643	386
Islam	309	82	321	343
Economics	74	304	1153	2227
English Language	36	16	184	35
Physics	130	236	397	557
Agriculture	250	203	4611	1231
Painting	32	64	431	110
English literature	96	206	1071	26
U. S. History	211	296	7714	283
TOTAL	1282	1769	17701	5516
AVE	128.2	176.9	1770.1	551.6

3. How is the hierarchical depth in DDC and subject directories?

Table 2 shows that Psychology and English language have the same hierarchical depth in DDC and The Open Directory. Painting is the only subject that has the same hierarchical depth in all directories. The selected subject directories don't have exactly the same hierarchical depth and have little in common with DDC. The hierarchical depth of Looksmart is the highest among the others and this is because of the expansion of subject directories through the names of countries, people, plants, etc. Table 2 also shows that DDC, in

comparison to Yahoo and The Open Directory, has a deeper hierarchical structure but the differences among them are negligible.

Table 2: The average of the hierarchical depth in DDC and subject directories.

Subjects	Yahoo	The Open Directory	Looksmart	DDC
Library & Information Science	4	8	7	5
Psychology	5	5	8	5
Islam	5	4	5	6
Economics	5	4	8	7
English Language	4	2	5	2
Physics	4	4	8	6
Agriculture	5	5	14	6
Painting	4	4	4	6
English literature	4	7	8	3
U. S. History	9	4	13	5
AVE	4.9	4.7	8	5.1

4. What is the frequency of subdivisions in the hierarchical levels?

Level 2 has the maximum frequency of subdivisions in Yahoo and The Open Directory (Table 3 and Figure 2). In Looksmart, level 3 has the maximum frequency of subdivisions (Table 3). But in DDC, level 4 has the maximum frequency of subdivisions (Table 4 and Figure 4). Table 3 shows that level 2 in subject directories has the highest frequency of subdivisions and this is different from what is in DDC. The frequency of subdivisions in DDC is not very different in all levels and shows a rather evenly distribution of subdivisions among all levels. They are 3, 3 and 4 in level 2, level 3 and level 4, respectively.

Table 3: The frequency of subdivisions in the hierarchical levels of subject directories.

Subject directories	Level 2	Level 3	Level 4	Level 6
Yahoo	7	2	1	0
The Open Directory	9	1	0	0
Looksmart	1	5	3	1
Total	17	8	4	1

Table 4: The frequency of subdivisions in the hierarchical levels of DDC.

	Level 2	Level 3	Level 4
DDC	3	3	4

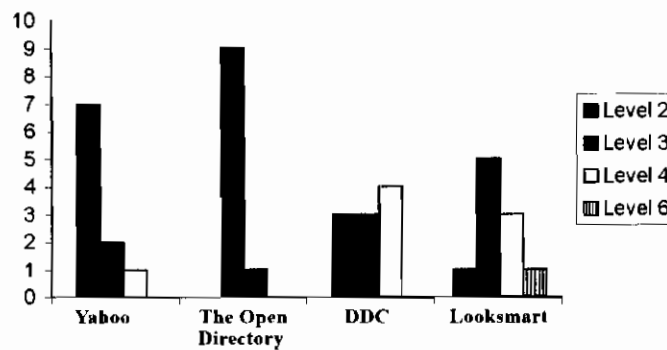


Figure 4: The frequency of subdivisions in the hierarchical levels of subject directories and DDC.

DISCUSSIONS

Data analysis shows that subject directories like Yahoo, The Open Directory and Looksmart do not possess exactly the same hierarchical structures and also they have little in common with DDC. These subject directories don't follow principles for organizing information as we can observe in library classifications and there is no global scheme for organizing Internet resources through directories. DDC, however, is a standard and universal classification scheme.

The hierarchical structure of Looksmart is broader than Yahoo, The Open Directory and DDC and this is because of its expansion through the names of countries, people, plants, etc. Although Yahoo and The Open Directory use the same pattern of expansion, they are not using it excessively. These subdivisions can be made in DDC by using its tables, such as tables of standard subdivisions and geographic areas and persons. The hierarchical structure of Looksmart is deeper than Yahoo, The Open Directory and DDC. As a result, it is more difficult to access subjects in Looksmarts than other directories and DDC.

The number of subdivisions at different levels of hierarchical structure is very different in Yahoo, The Open Directory and Looksmart. However, the numbers of subdivisions at different levels in DDC are very close to one another. This is because DDC uses parallel numbers and has, as one of its principles, to keep equal the number of subdivisions under each subject. DDC has other potentials, such as its general nature. Its general nature makes it a suitable classification scheme for organizing the Internet resources [5] because the Internet is a public network and includes any subject [3].

CONCLUSIONS

Findings reveal that the hierarchical positions and the frequency of subdivisions are different in DDC and subject directories. Yahoo, The Open Directory and Looksmart don't follow a unique, standard and global scheme to arrange their information. However, directories are effective mechanisms for classification of Internet resources. Users are able

to browse through the subject directories and retrieve related items. The hierarchical structure of Looksmart is broader and deeper than Yahoo, The Open Directory and DDC. This feature makes retrieval of information difficult by navigation through Looksmart subject categories and subdivisions. In comparison to subject directories, which were subjects of this study, DDC has an even distribution of subdivisions in its different hierarchical levels. This study approves the findings of the previous studies and shows that DDC can effectively be used in organizing and retrieving Internet resources.

REFERENCES

- [1] Balas, J., "Dewey and the Net." *Computers in Libraries*, Vol. 16, No. 1, pp. 46-47, 1996.
- [2] Botley, S., "Cataloging and Classification of Electronic Resources: Old Skills in a New Environment." *Aslib Managing Information*, pp. 31-32., 1999.
- [3] Mckiernan, G., "Beyond Bookmarks: a Review of Frameworks, Features and Functionalities of Schemes for Organizing the Web." *Internet Reference Services Quarterly*, Vol. 3, No. 1, pp. 69-82, 1998.
- [4] Rezaie Rad, M. and Abdorahmani, R., "Searching Internet Resources Through Classification or Yahoo Directory with Applied Vision on Army Sciences." in: *The First Information Conference of Army Forces*. Tehran: Malek-e-Ashtar Industrial University, pp. 857-876, 2001.
- [5] Vizine-Goetz, D., "Using Library Classification Schemes for Internet Resources." Available: at < <http://staff.oclc.org/~vizine/Intercat/vizin-goetz.htm>>, 1996.
- [6] Vizine-Goetz, D., "Classification Schemes for Internet Resources Revisited." Available: at < [http://staff.oclc.org/~vizine/JIC/v5n42002/classification schemes Revisited.doc](http://staff.oclc.org/~vizine/JIC/v5n42002/classification_schemes_Revisited.doc)>, 2002.
- [7] Wheatley, A., "Subject Trees on the Internet: A New Role for Bibliographic Classification?" *Journal of Internet Cataloging*, Vol. 2, Nos. 3-4, pp. 115-141, 2000.
- [8] Woodward, J. "Cataloging and Classifying Information on the Internet." *Annual Review of Information Science and Technology*, Vol. 3, No. 1, pp. 189-220, 1996.