# Models for Classifying Internet Resources

Chaim Zins

Department of Human Sciences, University of Haifa

Dr. Chaim Zins is an information scientist, a faculty member in the Department of Human Sciences in the University of Haifa. Dr. Zins' current research is focused on knowledge organization, knowledge management, information and referral services, and information retrieval.

Chaim Zins. (2002). **Models for Classifying Internet Resources.** *Knowledge Organization, 29(1).* 20-28. 25 refs.

**ABSTRACT:** Designing systematic access to Internet resources is a major item on the agenda of researchers and practitioners in the field of information science, and is the focus of this study. A critical analysis of classification schemes used in major portals and Web classified directories exposes inconsistencies in the way they classify Internet resources. The inconsistencies indicate that the developers fail to differentiate the various classificatory models, and are unaware of their different rationales. The study establishes eight classificatory models for resources available to Internet users. Internet resources can be classified by subjects, objects, applications, users, locations, reference sources, media, and languages. The first five models are content-related; namely they characterize the content of the resource. The other three models are format-related; namely they characterize the format of the resource or its technological infrastructure. The study identifies and formulates the eight classificatory models, analyzes their rationales, and discusses alternative ways to combine them in a faceted integrated classification scheme.

## Introduction

This study focuses on a major issue that researchers and practitioners in the field of information science currently face, namely, designing systematic access to Internet resources (IR). A brief review of Web portals and classified directories exposes different approaches to classifying the resources. These are reflected in the terminology, the number of categories, and the structures of the various classification schemes. Differences among resources have usually originated from diverse disciplinary, ideological, professional, cultural, and commercial considerations; contradictory assumptions regarding prospective users; and explicit and hidden interests that direct the developers. This divergence can explain, for instance, cultural differences between American and European directories regarding the "kicking ball game". In Europe (and the rest of the world) this game is called "football", not "soccer".

Although classification schemes are primarily evaluated on the basis of their practicality and their contribution to facilitating information retrieval, they are expected to be consistent and systematic. Michèle Hudon (2000) found inconsistencies in Web classification schemes. Inconsistencies present on the same Web page are problematic. For example, the main menu in *MSN* homepage (http://www.msn.com) included, as of November 22, 2001, a mixture of categories that reflect different classificatory criteria. Among them are the categories *Entertainment* and *Sports*, which represent two distinct subject domains, and the categories *Kids* and *Women*, which represent two distinct groups of prospective users. It also includes the categories *Careers*, *Games*, and *Travel*, which stand for users' applications, and the category *Music and Radio*, which indicates the media of the linked resources.

Looking at the homepage of dmoz: o*pen directory* (http://dmoz.org), one can trace at least five different classificatory models: by subject (e.g., *business*, *health*), by application (*shopping*), by categories of users (*kids and teens*), by location (e.g., *Canada*, *UK*, *US*), and by language (e.g., *Deutsch*, *Espanol*).

These inconsistencies might indicate that the developers fail to differentiate the various classificatory models, and are not aware of the rationale, strength, and limitations of each model. While each model has

its own rationale, different models should be implemented in the same classification scheme systematically, to enable users to utilize them efficiently. Reasonable ways exist to implement different models in the same classification scheme, but first we need to identify them. This study was aimed at identifying classificatory models used by major portals and classified directories. It identified eight models, demonstrated them, formulated their guiding principles, analyzed their rationales, and discussed optional ways to utilize them in an integrated classification scheme.

## Internet Classification Schemes

The advantages of classification schemes for organizing IR and for facilitating efficient searching have been discussed in the literature (e.g., Dodd, 1996; Gullikson, Blades, Bragdon, McKibbon, Sparling, & Toms, 1999; Koch, Brummer, Day, Hiom, Peereboom, Poulter, & Worsfold, 1997; Kwasnik, 1999; MacLennan, 2000; Mai-Chan, 1995; McIlwaine & Williamson, 1999; Molholt, 1995; Palmquist & Sokoll, 1998; Pollitt, Tinker & Braekevelt, 1998; Saeed and Chaudry (2001); Tinker, Pollitt, O'Brien & Braekevelt, 1999; Vizine-Goetz, 1996, 1998; Zins, 2000). Note that most of these references either describe ongoing projects or explore the implementation of library classification schemes, such as the Library of Congress Classification (LCC), the Dewey Decimal Classification (DDC), and the Universal Decimal Classification (UDC) in Internet projects. The advantages of classifying IR have also been acknowledged in practice by the increasing number of projects aimed at utilizing library and academic-based classification schemes (listed in Koch et al., 1997; McKiernan, 1998, 1999, 2000; Saeed & Chaudry , 2000; Woodward, 1996) and by numerous commercial Internet gateways, portals, and classified directories.

Classification schemes of IR are implemented in Web directories and classified lists for two interrelated purposes: facilitating an efficient information search and organizing the resources in a manageable way. Web directories and classified lists are evaluated in light of their contribution to facilitate efficient searching. The evaluation is usually based on three criteria: recall, precision, and cost effectiveness. Koch et al. (1997) found that classification schemes facilitate browsing and navigation through large collections, enable the broadening and/or narrowing of searches, provide the contexts to searches, and offer other advantages. They also found it useful for inexperienced searchers, and when the searcher is not familiar with the specific subject matter. Zins (2000) found that browsing through structured lists is especially useful in a well-defined subject context.

Organizing the resources in a manageable structure is good for developers and webmasters, as well as for Internet end-users. Developers and Website information managers use the classification scheme to organize the linked resources in a way that stresses thematic relations among them and to guide users to selected resources that meet the pre-selected thematic criteria. Organizing the resources in a manageable way is useful for more than facilitating efficient searching and retrieval. Developers and webmasters can utilize classification schemes for evaluating the overall coverage of the information domain, for evaluating the content of specific resources, and for reducing information overload by exposing duplication.

Internet users too can benefit from structured interfaces. A structured classification scheme helps users to overcome the perplexing effect of the chaotic nature of the Internet by providing a cognitive model of the information domain (see, for example, Kwasnik & Liu, 2000; Newton, 2000).

However, very often a cognitive model of the information domain can be misleading, since it may be based on a biased, illogical, or inconsistent scheme. Predefined schemes enable designers to manipulate users by stressing the thematic relations that best suit their personal biases, commercial interests, and ideological preferences. We cannot prevent these inherent limitations, but we can avoid designing ostensibly logical schemes based on unsuitable classificatory models or on inconsistent and confusing implementation of suitable models.

First we need to identify optional models for classifying IR in Internet search tools. Koch et al. (1997) list seven types of classifications: by subject, by language, by geography, by creating/supporting body, by user environment, by structure (enumerative or faceted), and methodology (a priori or empirical). Hudon (2000) lists four categories: disciplines or themes (e.g., education, leisure), forms of presentation (e.g., reference works, dictionaries), potential users, and geographic names (e.g., California).

Let us analyze the phenomenon to see if we can formulate a phenomenological-based typology of Internet classificatory models. The classificatory models in Internet search tools aim at linking users with remote information resources. There are three common places that can affect the relevance of the needed information: the user, the remote resource, and the interface (i.e., the Internet gateway).

The user works in a specific environment and location, is bound to a specific organization, and is aimed at accomplishing a specific application. This sets the groundwork for four optional classificatory models organized: (1) by users, (2) by applications, (3) by locations or environments, and (4) by organizations (e.g., hospital, library, school, university). The remote resource is of a specific type (e.g., a book, an encyclopedia article), is physically located at a specific site, is accessed by a specific URL, has a specific content, and utilizes a specific technology (e.g., DVD, WWW), specific media (e.g., graphics, texts), and a specific language. Very often it represents a specific real-world object (e.g., celebrities, movies), is related to specific locations and environments (e.g., NYC), and was added to the Internet at a specific time. These set the groundwork for ten optional classificatory models: (1) by types of information resources, (2) by servers' physical locations, (3) by domain names, which indicate the type of organization (e.g., commercial, educational, government) and the geographic location (e.g., CA, IL, UK), (4) by content or subjects, (5) by technology, (6) by media, (7) by languages, (8) by types of real-world objects, (9) by environments or geographic locations, and (10) by publication date.

The user interface is dependent on the developers' hidden and explicit agenda (i.e., business interests, ideological biases, professional guidelines) and on the technology used to facilitate the accessibility of remote resources. Some innovative technologies enable portal developers to manipulate the information available to users by adjusting the links in the accessed Web sites to refer users to pre-selected sites. This sets the groundwork for additional classificatory criteria based on developers' business interests and ideological preferences.

As we see, more than fifteen optional models exist for classifying IR, although some of them are interrelated. Therefore, we need to identify which of them are currently utilized in Internet gateways.

## Methodology

The study of Internet classificatory models followed a qualitative research methodology known as "grounded theory methodology". This is a general research methodology for developing theory, in this case formulating classificatory models, that are grounded in data, systematically gathered, and analyzed (Strauss & Corbin, 1994). The data for grounding the study were classification schemes used in homepages of nine sizable and highly popular general Internet portals. The Web sites were accessed on November 22, 2001.

The nine Internet gateways were of three types: general commercial portals, librarians' selections of recommended sites, and Web versions of library classification schemes. The first group was composed of five commercial portals: *AOL* (http://www.aol.com), *dmoz: open directory* (http://dmoz.org), *LookSmart (*http://www.looksmart.com), *MSN* (http://www.msn.com), and *Yahoo* (http://www.yahoo.com). The second group, librarians' selections of recommended resources, consisted of the *Librarians' Index to the Internet* project (http://www.lii.org). The third group contained three projects that utilize the three leading universal library classification schemes, the *Library of Congress Classification* (LCC), the *Dewey Decimal Classification* (DCC), and the *Universal Decimal Classification* (UDC) schemes. The three Internet projects are *ICRC: The Internet Collegiate Reference Collection: Contents by Library of Congress Classification* (http://library.bloomu.edu/reference/icrc/lc.php), which uses LCC, *CyberDewey* (http://www.anthus.com/CyberDewey/CyberDewey.html), which utilizes DDC, and *NISS: Directory of Networked Resources* (http://www.niss.ac.uk/subject/index.html), which uses UDC.

The classification schemes presented in the homepages of these nine selected portals were textually analyzed. Each term in the classification schemes, that is, each category and sub-category, was examined and was attributed to the most suitable pre-defined classificatory models. Three factors were taken into consideration for the attribution of the terms to specific models. These are the conceptual framework (i.e., the meaning of the term and its logical relations with other terms), the context (i.e., the term's position in the scheme), and the nature of the linked Internet resources.

In many cases the terminology could possibly stand for different classificatory criteria, requiring a supplementary analysis of the linked Web page . The *Animals* category in *Yahoo*'s classification scheme exemplifies such cases. At first glance, this category seems to refer users to a group of objects (i.e., taxonomy of animals). In fact, it stands for a subject domain, that is, it covers the field of "zoology". Another example was found in *MSN*. The *Women* category, at first glance, seems to stand for a subject domain (i.e., the field of women's studies). In fact, it includes resources, on various topics, designed for women. Therefore, in this case the *Women* category indicates

that the classificatory criterion is the prospective users rather than the subject of the linked resources: it contains resources for women rather than resources about women. Another example is the *People* category. In the *Librarians' Index to the Internet*, it stands for an object domain by including sites that represent a real-world object - people. But in *Open Directory* and *Yahoo* it stands for a subject domain; and in *MSN* and *AOL* it stands for an application domain, namely searching for people.

Portals were analyzed at the hierarchies' top level, which is the most significant part of the scheme. It is the first part that the user encounters when accessing the portal. In addition, it has the broadest coverage of the information domain. Furthermore, it is assumed that if the coverage of the subject domain is exhaustive, and if one takes the analysis down the hierarchical levels, one will eventually find sub-classes that are attributed to nearly all the classificatory models. Therefore, the analysis was primarily focused on the schemes' top level, while the second-level analysis was done only when it was needed to reach a clear conclusion regarding the right model. Note, however, that all the examples listed in this paper were verified by a review of the contents of the linked Web pages.

## Findings

### *Classificatory Models*

The analysis of the nine portals' homepages traced at least eight models for classifying IR. Internet information resources can be organized by (1) their subjects, (2) the real-world objects they represent, (3) the applications they are designed to meet, (4) the prospective end users, (5) the locations they are associated with, (6) their types as reference sources, (7) the media in which they are presented, and (8) the languages of the written texts. The eight models are shown in Figure 1.

***The subject model.*** The subject model is presented by categories such as *Business* (*Open Directory*), *Education* (*Looksmart*), *Health* (*AOL*), *History* (*ICRC*), *Law* (*Librarians' Index*), *Philosophy* (*NISS*), *Religion* (*CyberDewey*), *Science* (*Yahoo*), and *Sports* (*MSN*), which characterize the subject of the linked resource. Note that the parentheses enclose only one relevant portal for each term. All nine explored portals utilize the subject model.

***The object model.*** The object-oriented model is reflected in the following categories: *Associations* (*ICRC*), *Museums* (*ICRC*), *Organizations* (*CyberDewey*), and *People* (*Librarians' Index*), which indicate the type of the real-world object represented by the linked resources. Five portals utilize the model: *CyberDewey*, *ICRC*, *Librarians' Index*, *Looksmart*, and *Yahoo*.

Internet resources can be organized by:
1. **Subjects**
   (*Arts, Business, Education...*).
2. **Objects**
   (*Organizations*, *People...*).
3. **Applications**
   (*Chats, Email, Shopping...*).
4. **Users**
   (*Kids, Seniors*, *Women...*).
5. **Locations**
   (*Canada*, *UK*, *US...*).
6. **Reference**
   (*Almanacs*, *Dictionaries, Maps...*).
7. **Media**
   (*Graphics*, *Pictures, Radio...*).
8. **Languages**
   (*Deutsch, Espanol, Francais...*).

*Figure 1. Eight models for Classifying Internet Resources.*

***The application model.*** The application model is presented by terms such as *Chat & People* (*AOL, MSN*), *Personals* (*Looksmart*), *Searching the Internet* (*Librarians' Index*), *Shopping* (*Open Directory*), and *Travel* (*Looksmart*), which refer the user to resources that are related to the specified applications. Six portals use the model: *AOL*, *Librarians' Index*, *Looksmart, MSN, Open Directory*, and *Yahoo*.

***The user model.*** The user model indicates the prospective end users of the linked resources. This is reflected in categories such as *Families* (*Librarians' Index*), *Kids & Teens* (*Open Directory*), *Professions* (*Looksmart*), *Seniors* (*Librarians' Index*), and *Women* (*MSN, AOL*). Generally, there are three types of prospective users: (1) Age groups (e.g., *Kids, Seniors*); (2) Professional groups (*Professions*), and (3) Interest groups. The interest groups are organized on the basis of a common interest such as gender (*Men & Women*) or social/human condition/problem (e.g., *Disabled*, *Gay* & *Lesbian*, and *Families*; these examples are taken from *Librarians' Index*). The six following portals use the model: *AOL*, *Librarians' Index, Looksmart, MSN, Open Directory,* and *Yahoo.*

***The location model.*** The location model is used for referring users to resources relevant to the indicated geographic or national locations by their content. It is usually presented in most portals by the *Regional* category (e.g., *Open Directory* & *Yahoo*), or by referring users to relevant directories, as in the case of *City*

*Directories, Global Directories* (*Looksmart*), and *Local Yahoos* (*Yahoo*). Five portals, *Librarians' Index, Looksmart, MSN, Open Directory*, and *Yahoo*, utilize the *Location* model.

***The reference model.*** The reference model is usually reflected in the *Reference* category (e.g., *Librarians' Index*, *Looksmart*, and *Open Directory*) and the *Generalities* category (*CyberDewey*, *NISS*). It is used for referring users to different reference sources such as almanacs (*ICRC*), dictionaries (*Librarians' Index)*, encyclopedias (*CyberDewey*), magazines (*CyberDewey*), and maps (*AOL, Open Directory*). *AOL, CyberDewey, ICRC, Librarians' Index, Looksmart, MSN, NISS, Open Directory,* and *Yahoo* use the model, a total of nine portals.

***The medium model.*** The medium model points out the medium by which the content and the messages of the specified resource are expressed or communicated. It is reflected in categories such as *Graphics, Pictures,* and *Radio* (*Librarians' Index*). The term "Medium" is used here in its broader sense, as by Marshal McLuhan in his book *The Medium is the Message* (McLuhan & Fiore, 1967). It is the electronic means or the electronic technology by which we communicate. In this sense the Internet itself is a medium. Clearly, the medium model, as formulated here, incorporates the two models, technology and media, that were identified earlier in this study into one model. Four portals, *Librarians' Index, MSN, Open Directory,* and *Yahoo*, use the model.

***The language model.*** The language model organizes resources by their text languages. *Open Directory* provides a good example of a portal that implements the model. It refers its users to lists of sites in a variety of languages, including *Deutsch, Espanol, Francais, Polska,* etc. The language model is used by four portals: *Librarians' Index*, *MSN, Open Directory, and Yahoo.*

***Rating.*** Based on the number of portals that utilize each model, the eight classificatory models are rated as follows (note that the numbers in the square brackets indicate the number of portals that use the model). The *subject* model [9] and the *reference* model [9] are the most common, being implemented by all nine portals. The *application* model [6] and the *user* model [6] are applied by six portals each. The *object* model [5] and the *location* model [5] are applied by five portals. Finally, the *medium* model [4], and the *language* model [4] are practiced by four portals. Figure 2 summarizes the findings.

| | Subject | Object | Application | User | Location | Reference | Medium | Language |
|---|---|---|---|---|---|---|---|---|
| AOL | X | | x | x | | x | | |
| CyberDewey | X | x | | | | x | | |
| ICRC | X | x | | | | x | | |
| Librarians' Index | X | x | x | x | X | x | x | x |
| Look Smart | X | x | x | x | X | x | | |
| MSN | X | | X | x | X | x | x | x |
| NISS | X | | | | | x | | |
| Open Directory | X | | X | x | X | x | x | x |
| Yahoo | X | x | X | x | X | x | x | x |

*Figure 2. The findings*

*Internet Gateways*

One portal, *NISS,* implements two models. Two portals, *CyberDewey* and *ICRC,* implement three models. One portal *AOL*, implements four models; one, *Looksmart,* implements six models; two, *Open Directory* and *MSN,* implement seven models; and two gateways, *Librarians' Index* and *Yahoo*, implement all eight models.

***AOL.*** *AOL* uses four models, namely, *subject* (e.g., *Entertainment*, *Health*), *application* (e.g., *Classifieds, Horoscopes, Weather*), *user* (e.g., *Women*), and *reference* (e.g., *Maps, White pages, Yellow pages*).

***CyberDewey.*** *CyberDewey* utilizes three models, namely, *subject* (e.g., *Literature*, *Religion, Technology*), *object* (e.g., *Organizations & Museums*), and *reference* (e.g., *Encyclopedias*, *Magazines*).

***Dmoz: Open Directory.*** *Open Directory* uses seven models. These are *subject* (e.g., *Arts, Business, Health*), *application* (*Shopping*), *user* (*Kids & Teens*), *location* (e.g., *Canada, UK, US*), *reference* (*Reference, Newspapers*), *medium* (*Newspapers*), and *language* (e.g., *Italiano, Japanese, Svenska*). Note that the category *Newspapers*, in this analysis, stands for two models, *reference* and *medium*, though it appears in the Web site

only once, as a sub-category of *News,* rather than as a sub-category of *Reference*. Evidently, according to the editors of *Open Directory*, newspapers are a medium for communicating news since they appear as a sub-category of *News.* However, they can also be viewed as reference resources, like books, magazines, and maps.

***ICRC.*** *ICRC* implements the *subject* (e.g., *History, Philosophy, Psychology, Religion*), the *object* (e.g., *Associations, Museums*), and the *reference* (e.g., *Almanacs, Encyclopedias, Indexes*) models.

***Librarians' Index.*** *Librarians' Index* utilizes all eight models: *subject* (e.g., *Arts*, *Business*, *Law*), *object* (e.g., *Associations, Nonprofit Organizations, People*), *application* (e.g., *E-mail, Jobs, Searching the Internet*), *user* (e.g., *Gay, Lesbian & Bisexual, Kids, Seniors*), *location* (*California*), *reference* (e.g., *Almanacs*, *Dictionaries*, *Phone Books*), *medium* (e.g., *Images: Clip Art, Graphics, Pictures, Radio*), and *language* (e.g., *English*, *Spanish*).

***Looksmart.*** There are six classificatory models in *Looksmart*: *subject* (e.g., *Computer Science*, *Education*), *object* (e.g., *Celebrities*, *Companies)*, *application* (e.g., *People & Chat*, *Shopping, Travel*), *user* (*Family*, *Kids*, *Professions*), *location* (*City Directories, Global Directories*), and *reference* (*Reference*).

***MSN.*** *MSN* implements the following seven models: *subject* (e.g., *Autos*, *Entertainment*, *Sports*), *application* (e.g., *Careers, Games, Travel*), *user* (*Kids, Women*), *location* (*e.g., Canada, City Guides, Mexico*), *reference* (*City Guides, Maps*), *medium* (*Radio*), and *language* (*Spanish*).

***NISS.*** *NISS* utilizes two models, *subject* (*Philosophy, Social Sciences, Technology*), and *reference* (*Generalities*).

***Yahoo.*** The ninth gateway, *Yahoo,* utilizes all eight models: *subject* (e.g., *Literature, Science*), *object* (*Clubs, Members*), *application* (e.g., *Career*, *Chat*, *Greetings*), *user* (*Kids*), *location* (*Local Yahoos*), *reference* (e.g., *Dictionaries*), *medium* (*Radio*), and *language* (*Chinese, Spanish*).

## Discussion

### *Rationales*

The term "model" is used throughout the paper in two distinct, though related, meanings: as a principle of division and as a scheme. Note that the eight models (meaning, principles of divisions) prepare the groundwork for eight different models (meaning, classification schemes). Each of the eight models (meaning, principles of divisions, and classification schemes) has its own rationale, which establishes its advantages and justifies its implementation.

***The subject model.*** The subject classification presents a knowledge structure of the information domain. It is based on the following rationale: a Web page is an item of information; items of information differ from each other; and are characterized by their subjects. The *subject* model is meant to organize the various Internet resources, as well as facilitate efficient information retrieval, by grouping the resources according to their common thematic attributes. It is the most fundamental model and is utilized by all nine explored portals. Note that all the major universal library classification schemes, including Ranganathan's faceted Colon Classification (Kaula, 1985), are subject models.

***The object model.*** The object classification presents a taxonomy, or a typology, depending on the structuring methodology, of the relevant real-world objects. It is based on the rationale that many Web pages represent objects in the real world (e.g., books, celebrities, cities, hospitals, newspapers). Therefore, object-oriented classifications are primarily aimed at assisting users to locate the Web sites of specific real-world objects.

One may argue that the reference, the medium, and the location classifications are three sub-sections of the object classification, since they organize three types of real-world objects. To be sure, the object classification is inclusive. It embraces reference resources, media-related resources, and geographic locations. Nevertheless, each of these three models has a unique rationale, which justifies its treatment as a distinct classificatory model.

Furthermore, one may argue that the *object* model should include all types of objects, real and virtual. This approach broadens the scope of the *object* classification scheme by including concepts and other types of knowledge representations. According to this approach, Web resources that represent abstract concepts, such as "knowledge organization", "love", and "democracy", can be classified in the *object* model as *general concepts*, as well as in the *subject* model.

***The application model.*** The application classification presents a taxonomy of Internet applications. Its rationale is the assumption that people use the Internet to accomplish a specific application, namely sending email, chatting with a remote person, searching for information, or participating in a professional forum. Therefore, classifying the resources according to the application they are likely to meet seems extremely useful.

***The user model.*** The user classification presents a taxonomy of Internet users. The grouping is based on identified common characteristics of prospective users. There are at least three criteria for characterizing users: by age (age groups), by profession and occupation (professional groups), and by focus of interest (interest groups). The model is based on the rationale that Internet resources are primarily aimed at meeting the needs of their prospective users. Therefore, it seems highly useful to group the resources according to the prospective users.

***The location model.*** The location classification presents the geographical division of the world. However, in practice it is useful to base the classification scheme on the world's political division into states, political entities, and regions. The model is based on the following rationale: Internet users live in the real world (as opposed to the virtual cyberspace), are bound to their real communities, and have interests related to real places. Evidently, the *location* model fulfills a significant role in assisting users to locate resources related to real places.

***The reference model.*** The *reference* classification presents a taxonomy of reference sources and search tools accessible through the Internet. The rationale of the model is that users very often need different reference sources such as books, desk reference sources (e.g., almanacs, dictionaries, encyclopedias), journals, or search engines to meet their information needs. The reference classification is aimed at facilitating direct access to Internet search tools and information reference sources. The fact that the model is used by all nine portals reflects the centrality of information searching as a dominant application of the Internet. Therefore, developers who utilize the model are required to anticipate the information needs of the prospective users and provide relevant tools to meet them.

***The medium model.*** The medium classification presents a taxonomy of the information technologies utilized through the Internet. The Internet is by its nature a technological arena. The model is intended to assist users to locate resources related to a specific technology. Yet in light of the ever-changing nature of information technologies, classifying resources by their technology very often tends to be outdated. It seems that the model is implemented mainly for locating resources related to recent innovations or for locating popular applications.

Note that the reference classification can be viewed as a sub-section of the medium classification, since reference resources can be viewed as media for communicating information. However, these two models have different rationales, so it seems reasonable to formulate two distinct classifications.

***The language model.*** The eighth classification, language, presents a taxonomy of languages used by Internet users. It seems to be very useful for classifying resources in multi-cultural environments common to English-speaking and non-English-speaking users, by providing non-English-speaking users with efficient access to non-English resources.

*Integrated Classification Schemes*

The eight models prepare the groundwork for possible integrated classification schemes. While implementing the models in a systematic scheme, developers of Web portals should relate to the relevant models, their order in the integrated scheme, and the type of integration.

The number of relevant models varies from one through eight. Designers can utilize only one model, some of them, or all the eight models. Utilizing only one of the models might be the most economical. I have observed that the *subject* model is implemented by all nine explored portals. So if a Web site's developer is required to base the categorization exclusively on one of the models, it will most likely be the *subject* model.

Utilizing some of the classificatory models in an integrated scheme enables developers to focus attention on the most significant perspectives. Note that the eight models are of two types, namely content-related and format-related. The first five models – *subject, object, application, user,* and *location* – are content-related: they depict the content of the resource. The last three models – *reference, medium*, and *language* – are format-related: they depict the format or the technological infrastructure of the related resource. It is difficult to determine a priori which models are significant for the prospective users. Still, the five content-related classificatory models, together with the reference model, seem to be very useful for facilitating efficient information retrieval.

Utilizing all the eight models enables developers to stress all the different perspectives while organizing the resources. Furthermore, all the classificatory models can be used as guidelines for formulating keywords and structured thesauri designed for use in metadata fields (e.g., *Keywords* and *Descriptions*) or in metadata forms (e.g., *the Dublin Core*). Still, utilizing all the models in an integrated scheme or a metadata form appears to be wasteful, since the models are not

mutually exclusive. Therefore, a further analysis should be made by the portals' developers to avoid unnecessary duplications.

As noted above, utilizing more than one model in an integrated faceted scheme obliges developers of Web portals to decide on the order of the integrated models and the type of integration. The order of the models in this paper – *subject, object, application, user, location, reference, medium*, and *language* – reflects my own preferences. The integrated scheme can be in the form of Figure 1 (see above), in which all of the classificatory models appear in the top level. The reader can see an example of such integration in the *Success* web site (http://www.success.co.il). Another form is set by Ranganathan's faceted Colon Classification (Kaula, 1985), in which the top level presents the key model and the other models are presented down the hierarchies, in the sub-classification levels.

*Research Agenda*

In the phenomenological analysis presented above, I found more than fifteen optional classificatory models for classifying IR. In this study, I found that only eight models are used in the nine explored Internet portals. By the time this paper is published, the reader may well find that the number of the implemented models is different. As noted above, each of the eight models has a unique rationale, which theoretically justifies its utilization. However, its practical justification depends on its actual contribution to meeting the information needs of the relevant prospective users, and, more significantly, on the model's cost effectiveness in terms of cost per recall and precision. This sets a research agenda for user studies as well as for design studies.

## Conclusion

The study establishes that at least eight classificatory models exist for classifying Internet resources: by subjects, by objects, by applications, by users, by locations, reference sources, media, and languages. The first five models are content-related as they characterize the content of the Internet resource. The other three models are format-related in that they characterize the format of the resource or its technological infrastructure. The eight models set the groundwork for possible integrated faceted classification schemes. Developers of portals can utilize all eight models, some of them, or only one. Still, utilizing the different classificatory models in portals' user interfaces requires development of the related systematic classification schemes and exploration of the different ways to integrate them. This sets a research agenda for portal developers. Nevertheless the structured schemes have to be efficient, and this sets a research agenda for experts in users' behavior.

## References

Dodd, D. G. (1996). Grass-Roots Cataloging and Classification: Food for Thought from World Wide Web Subject-Oriented Hierarchical Lists. *Library Resources & Technical Services*, 40 (3), 275-285.

Gullikson, S., Blades, R., Bragdon, M., McKibbon, S., Sparling, M., & Toms, E.G. (1999). The Impact of Information Architecture on Academic Web Site Usability. *The Electronic Library*, 17 (5), 293-304.

Hudon, M. (2000). Innovation and Tradition in Knowledge Organization Schemes on the Internet, or, Findings One's Way in the Virtual Library. In Beghtol, C., Howarth, L. & Williamson, N. J., eds. (2000). *Dynamism and Stability in Knowledge Organization: Proceedings of the Sixth International Isko Conference*. Wurzberg, Germany: Ergon Verlag.

Kaula, P. N. (1985). *A Treatise on Colon Classification*. New Delhi, India: Sterling Publishers.

Koch, T., Brummer, A. B., Day, M., Hiom, D., Peerebom, M., Poulter, A., & Worsfold, E. (1997). *The Role of Classification Scheme in Internet Resource Description and Discovery* {WWW document]. (accessed June 1, 2002), URL: http://www.ukoln.ac.uk/metadata/desire/classification

Kwasnik, B.H. (1999). The Role of Classification in Knowledge Representation and Discovery. *Library Trends*, 48(1), 22-47.

Kwasnik, B.H. & Liu, X. (2000). Classification Structures in Changing Environment of Active Commercial Websites: the Case of eBay.com. In Beghtol, C., Howarth, L. & Williamson, N. J., eds. (2000), *Dynamism and Stability in Knowledge Organization: Proceedings of the Sixth International Isko Conference*. Wurzberg, Germany: Ergon Verlag.

MacLennan, A. (2000). Classification and the Internet. In Marcella, R. & Maltby, A. (2000), *The Future of Classification*, Great Britain: Gower.

Mai-Chan, L. (1995). Classification, Present and Future. *Cataloging & Classification Quarterly*, 21 (2), 5-17.

McIlwaine, I. C. & Williamson. N. J. (1999). International Trends in Subject Analysis Research. *Knowledge Organization*, 26 (1), 23-29.

McKieran, G. (1998). Beyond Bookmarks: A Review of Frameworks, Features, and Functionalities of Schemes for Organizing the Web. *Internet Reference Services Quarterly*, 3(1), 69-77.

McKieran, G. (1999). Points of View: Conventional and 'Neo-Conventional' Access and Navigation in Digital Collection. *Journal of Internet Cataloging*, 2(1), 23-41.

McKiernan, G. (2000). *Beyond Bookmarks*: [WWW document] (accessed May 2000), URL: http://www.public.iastate.edu/~CYBERSTACKS/CTW.htm

McLuhan, M. & Fiore, Q. (1967). *The Medium is the Message. An Inventory of Effects*. New York: Bantam Books.

Molholt, P. (1995). Qualities of Classification Schemes for the Information Superhighway. *Cataloging & Classification Quarterly*, 21(2), 19-22.

Newton, R., (2000). Information Technology and New Directions. In R. Marcella& A. Maltby, eds. *The Future of Classification* (pp. 43-57). Great Britain: Gower

Palmquist, R. A. & Sokoll S. P. (1998). Visual Maps of the World Wide Web: Helping the User Find the Way. *The Reference Librarian*, 60, 49-60.

Pollitt, A. S., Tinker, A. J. & Braekevelt. (1998). Improving Access to Online Information Using Dynamic Faceted Classification. *Online Information*, 98 Proceedings, 17-21.

Saeed, H., & Chaudry, A.S. (2001). Potential of Bibliographic Tools to Organize Knowledge on the Internet: The Use of Dewey Decimal Classification Scheme for Organizing Web Based Information Resources. *Knowledge Organization*, 26 (2), 80-96.

Strauss, A., & Corbin, J. (1994). Grounded Theory Methodology. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research* (pp. 273-285). Thousand Oaks, California: Sage.

Tinker, A. J., Pollitt, A. S., O'Brien, A. & Braekevelt, P.A. (1999). The Dewey Decimal Classification and the Transition from Physical to Electronic Knowledge Organization. *Knowledge Organization*, 26 (2), 80-96.

Vizine-Goetz, D. (1996). *Using Library Classification Schemes for Internet Resources.* [WWW document] (accessed June 1, 2002), URL: http://staff.oclc.org/~vizine/Intercat/vizine-goetz.htm

Vizine-Goetz, D. (1998). Dewey as an Internet Subject Guide. In W. Mustafa el Hadi, J. Maniez, & S.A. Pollitt, (eds.). *Structures and Relations in Knowledge Organization: Proceedings of the 5th International ISKO Conference*. Wurzberg, Germany: Ergon Verlag. [Abstract retrieved June 1 2002 from http://staff.oclc.org/~vizine/isko/vizine-goetz_isko5.htm 2002, June 1]

Woodward, J. (1996). Cataloging and Classifying Information Resources on the Internet. *Annual Review of Information Science and Technology*, 31, 189-220.

Zins, C. (2000). Success, A Structured Search Strategy: Rationale, Principles, and Implications. *Journal of the American Society for Information Science*, 51 (13), 1232-1247.