Abbie Anderson

L505: Organization and representation of knowledge and information

Spring, 2002

Instructor: Erika Johnason

Final Exam

5. "Organizing Information" from Rosenfeld and Morville's

<u>Information Architecture for the World Wide Web</u>

1) indicative abstract

Rosenfeld, R. and P. Morville. (1998). "Organizing Information". Chapter 3 in *Information Architecture for the World Wide Web* (22-46). New York: O'Reilly.

Describes challenges and strategies for structuring web sites and intranets. Promotes production of cohesive systems that aid navigation, meet user needs, and suit the nature of both content and use of the site. Provides general principles and specific examples to guide designers. –ama 4/20/02

2) list of descriptors from ASIS thesaurus (no more than 5)

- ❑ arrangement
- ❑ access to resources
- ❑ World Wide Web
- ❑ navigation (information retrieval)
- ❑ organization of information

3) subject headings from LCSH (no more than 5)

- ❑ Information organization
- ❑ System design
- ❑ Web sites--Design
- ❑ Electronic information resource searching
- ❑ Mental representation

4) critical essay

This chapter could almost be used as an introductory or overview reading for L505—making it, of course, a fitting close to our work as well. The

ambiguities of language and of subjective meanings; mental models; organization schemes; classification; controlled vocabulary—they're all here. We're only missing representation; more in-depth, library-oriented treatment of indexing and thesauri; the indexing of images; and metadata. Of course, the presentation is specialized, focused on information design for the web. The authors do refer to librarians several times, however (e.g., "we're all becoming librarians," p. 23; assigning items to categories "is a painstaking process that only a librarian could love," p. 31), thus reinforcing Erika's point in our 4/18/02 class session that librarianship really is all about information architecture.

As Rosenfeld and Morville put it in their opening statement for this chapter, "Our understanding of the world is largely determined by our ability to organize information." In the second paragraph they add, "We organize to understand, to explain, and to control." Finally, in the third paragraph, they affirm, "As information architects, we organize information so that people can find the right answers to their questions. … Our aim is to apply organization and labeling systems that make sense to users." Taken together, these statements neatly sum up first the material we covered at the beginning of L505, and then the reason we study such things in the first place.

Any organizational system is going to be based on *somebody's* mental model, someone's understanding of the world and its perceived constituents. This can be as subjective and individual as my idiosyncratic style/period organization of my CDs, as conventional and unreflective as the traditional sections of a cookbook, or as formal and consensus-driven as the biological classification of species. We learned in L505 that human beings 1) rely on mental models or schemes in order to interpret the environment and make decisions about how to act; 2) use mental models or frameworks to integrate new information into what is already known. We also learned that communicating our understandings of the world to someone else is far from guaranteed. The subjective nature of understanding and the ambiguous nature of language join forces against clear, direct, assured communication. Our readings for sessions three through five in this course revolved around these

same issues. In their section on "Organizational Challenges", Rosenfeld and Morville give a cogent exposition of material covered by such scholars as Lakoff & Johnson, Zerubavel, Norman, Rumelhart and Wurman in the required readings for L505.

The combined ideas that 1) people need frameworks (and will invent them if necessary, often in ways that a designer might not have predicted), and 2) people (and languages, and minds) are idiosyncratic, form a subtext for most of the authors' points in this chapter about what is important for structuring information. An organizational system must satisfy not just the standards and expectations of the designer, but those of the people to whom it will be presented for use. Those people may be approaching the system for very different reasons and with very different understandings of how to go about the task at hand—different from the designer, and different from each other. As several of our early readings in L505 stressed, the best way to do this—the *only* way to do this effectively—is to know your users and their needs. Happily, the authors manage to make these points without framing users as "the problem", but rather as the ones being served by the system, whom the system should fit.

Rosenfeld and Morville continue their voyage through L505 as they discuss "Organization Schemes" and "Organization Structures", entering the territories of categorization and classification that we covered in class sessions six through eight. While the authors do not provide a librarian's introduction to these issues (with no need to mention, for instance, the Library of Congress or Ranganathan), they do once again establish the principles involved with clarity and aplomb. Since they are not bound by the requirements or the history of academic information science, they are free to present these issues without delving into the controversies that necessarily took so much of our time in L505. Their discussion of how people seek information and how they navigate the system also brought to mind material covered in L524 and L503.

The authors present hierarchical structure as simple and familiar to people, and recommend that for most purposes, a hierarchy should probably form at least the outer framework to help people navigate your site—while also

stressing that different sections of the site should probably take advantage of different, more relational or data-oriented approaches. Their exposition of the strengths and weaknesses of the different kinds of structures was extremely useful, and helped to clarify some of our L505 discussions on a practical level. Their diagramming of "narrow and deep" vs. "broad and shallow", both visually and verbally, was especially valuable (for once a visual example worked for me!).

Within a discussion of database structure, the authors present controlled vocabulary as a tool for helping people browse and search. They advocate what amounts to building a thesaurus for the material at hand, and along the way touch on many of the issues we covered in our class sessions nine, eleven, twelve and thirteen. Controlled vocabulary offers a consistency in approach and certainty of language use that aids both the designer and the user— assuming the vocabulary is easily accessed, easily understood, and easily applied.

While Rosenfeld and Morville's aims are specialized (as indicated by their title!), the clarity and pragmatism with which they present and inter-relate many of the issues covered in L505 is most welcome at the end of this semester. It is not surprising that this book is considered the "bible" for information architecture and large-scale web design. Their commitment both to the needs of the user and the integrity of site content is impressive. While obviously one couldn't use a book like this as a textbook for L505, since it has no reason to cover the academic issues central to understanding the history of these ideas in librarianship, it performs admirably as a complement to the work we have done this semester.

4. Faceted classification for grocery store products

<u>The items:</u>

- ❏ A jar of sliced, roasted red peppers
- ❏ Chopped celery in cellophane wrapping
- ❏ Diced peaches (no sugar added) in a can
- ❏ Gallon jug of orange juice (no pulp)
- ❏ Ten pound sack of potatoes
- ❏ Frozen strawberries in heavy syrup (sugar added) in a paper carton
- ❏ Frozen corn in a plastic bag

<u>The facets:</u>

- ❏ *Products*
- o Sub-facets: **Type** and **Kind**
- o May sub-divide with intermediary level of category (e.g. Produce, Household, etc.; but not recommended: forces pre-coordinate groupings)
- ❏ **Treatments**
- ❏ **Features** Added/Excluded
- ❏ *Packaging*
- o Sub-facets: **Forms** and **Materials**
- ❏ **Avoirdupois** (size/weight/volume)

<u>Facet Schedule:</u>

Products

  For mixed kinds, indicate with Mi, constituents in parentheses, hyphenated; e.g., VeMi(Ca-Co-Pe). Varieties (such as Valencia orange) indicated with hyphen. [Actual scheme would include separate schedule for varieties]. Brands indicated by colon after variety [separate schedule for brands], e.g. FrOr-v:Su for Sunkist Valencia orange.

  Types

Kinds

Fr. Fruit

       Or. Oranges

       (Pa. Pears)

       Pe. Peaches

       St. Strawberries

Ve. Vegetables

       (Ca. Carrots)

       Ce. Celery

       Co. Corn

       (Pe. Peas)

       Po. Potatoes

       Pp-r. Peppers-red

       (tomato would be treated as a vegetable in the grocery context)

Treatments

Combine with hyphen, e.g. Fr-Ch; action done to whole product listed before actions which create subdivisions of product.

       Ch. Chopped

       Di. Diced

       Fr. Frozen

       Ju. Juice

       Ro. Roasted

       Sl. Sliced

Features

       Su. Sugar added

       Su0. No sugar added

       (Pu. Pulp)

       Pu0. No Pulp

Packaging

Where material can be assumed, as in "(glass) jar", Material sub-facet may be omitted.

Forms

Bg. Bag or Sack

Bt. Bottle or Jug

(Bx. Box)

Cn. Can

Ct. Carton

Ja. Jar

Tr. Tray

Wr. Wrap or wrapping

Materials

Gl. Glass

Me. Metal or foil

Pa. Paper

Pl. Plastic (incl. Cellophane)

(St. Styrofoam; compound TrSt-WrPl =styrofoam tray with plastic wrap)

Avoirdupois

Number plus a dash plus conventional abbreviation for unit of measure (e.g., 15-oz.)

Notation:

Periods between facets; hyphens for combined features within facets or sub-facets. No period at end of row.

Product elements (Type, Variety, Treatment, Features) on top row, Packaging elements (Form, Material, Avoirdupois) on second row of label. Notation may also be made on single row with a slash between Product and Packaging elements.

Facets which do not apply or are not known may be omitted; omission indicated by a period as placeholder. Notation enhanced by use of zero with slash (v. the letter O).

<u>Notation Order:</u>

TypeKind.Treatment.Features

FormsMaterials.Avoirdupois

Example 1: jar of sliced roasted red peppers (size not given)

VePp-r.Ro-Sl.

Ja.

Example 2: Gallon jug of orange juice (no pulp)

FrOr.Ju.Pu0

BtPl.1-gal

Example 3: Ten pound sack of potatoes

VePo../BgPa.10-lb

<u>Explanatory Notes:</u>

This classification aims to produce the hospitality in array required by a grocery store, which is always adding and removing products, as well as the clarity necessary for quick interpretation of notations. Letter-based notation was chosen to support these aims, allowing new products to be inserted easily into the schedules in a sensible manner. Letters also function as mnemonic aids, relating to the first two letters of the word represented (or the first letter followed by the most distinctive letter), thus reinforcing the ideal of sense-making in the system. Alphabetic order in the schedules also provides for ease of product look-up. In the database, products can be sorted or retrieved by packaging or by avoirdupois as easily as by product type or kind. The faceted scheme lends itself well to flexible manipulation in relational databases.

For labels on products, the two-line notation is preferred for ease of reading. This helps to reduce the impact of one of the main drawbacks of faceted schemes, the length of notation. A slash between the Product and Packaging facet groups serves the same "visual sorting" function when the store may prefer to place notations on a single line, as on file folders and in reports. Items may be assigned numerical bar codes separate from the faceted classification. Like an item's price, the individual bar code does not qualify as a "facet" to classify, but nevertheless has an important place in the database as well as on the label. However, one cannot look at a bar code and know immediately what item it represents, while the faceted notation is designed for just such interpretation.

Typical English word order has not been followed for notation order. In fact, the notation preferred for this classification follows almost the exact reverse of spoken English, as a "ten-pound bag of potatoes" becomes (translated from the notation) a "vegetable potato bag paper 10-lb." Although in some respects English word order would have made the notation easier to read, it would have imposed inconvenient structure on notation for clarity of classification. While we speak our modifiers first in English, the subject (potatoes) is the fact we need to know first when dealing with products in the grocery store.

Of course individual stores are free to devise notational orders which best suit their operations. In an actual formulation of this classification, lead-in terms in the schedule would clearly indicate the system's vocabulary, such as "bag" for "sack", or the combined facets "plastic wrap" for "cellophane".

1. Physical, conceptual and contextual surrogates

The human animal requires representations in order to communicate and manipulate ideas. Language is perhaps our most familiar and most necessary representation; writing, as a representation of language, one of our most important means of objectifying thought. A representation is something that stands in or stands for something else, the equivalent of or correspondent to something else. A representation, then, is a surrogate. Surrogates perform specific roles in information science for the referencing of documents, the ideas they contain, and the subject areas to which they relate.

Surrogates can be physical, conceptual, or contextual. A physical surrogate represents the characteristics of an item or document as a concrete entity. A conceptual surrogate represents the intellectual content or meaning of the item. A contextual surrogate places the content of the item in place against a specific background, normally a field of study or academic discipline. All three kinds of surrogates are used to describe the item, but each one plays a different kind of role in making the document accessible for retrieval.

A physical surrogate records data about the document as a physical object. To do so accurately and consistently it must be precise, expressed in a controlled format governed by a specific syntax, so that each surrogate presents its referent in the same way as the next. The meaning of the physical surrogate is contained within the strictures of its format. A good example of a physical surrogate is the bibliographic record—the catalog card of long library tradition, now translated as an OPAC entry with its various fields such as author, title, publisher, date of publication, subject headings, and shelf location. Physical surrogates tell us what we need to know about an item as a discrete, complete package, including where to find it and how it is classified. Physical surrogates, however, tell us little or nothing about what a document means or how it relates to broader areas of study. They may include clues to content or to a broader context via the items located in the same area or grouped under the

same subject heading, but those "clues" are there by inference only. Physical surrogates are self-contained: they refer only to the item, neither inside it nor outside it. Like lines in a telephone book, they give us names and places and means of contact, but not personalities or relationships (beyond the resemblance of family names or proximity of location).

Conceptual surrogates point not to the item itself, but to the ideas it contains. They rely on controlled vocabulary to meaningfully and consistently represent those ideas. They are concerned not with the entity itself, but with what it treats or what it means. They can be applied to an item to indicate its intellectual content. Subject headings and indexing terms are good examples of conceptual surrogates. A physical surrogate can include a conceptual surrogate (such as the subject heading in a bibliographic record), but not vice versa. A conceptual surrogate provides labels for ideas, but not for items as such. Once connected to a document, a conceptual surrogate can tell you about the ground covered by the item, but not where it is now or how it got there. Conceptual surrogates provide access to ideas within documents.

Contextual surrogates are the broadest type of representation, pointing neither to the document itself nor to specific concepts within it, but to the place of the document in its field of study. Of these three types of surrogates, they are least concerned with strict format, structure, syntax or vocabulary. Abstracts are good examples of contextual surrogates. The aim of the contextual surrogate is to indicate the import of the item—what it means not in terms of single concepts, but in terms of argument. Like a conceptual surrogate it is concerned with content, but not with specific ideas as individual units to be labeled; rather, it can be seen as a narrative of what the document is about. The aim of the contextual surrogate is to describe the thought or work of the document in such a way that the user can determine whether or not the document is applicable to his or her interests and also its value in it discipline. It will not categorize the document's content, however, nor will it tell you much about the item as an entity.

All three of these surrogates are useful for information retrieval, but in different ways. Physical surrogates support known item searches. The phone book, for instance, will let me look up my doctor's phone number so long as I know his name. A set of physical surrogates may allow for browsing, depending on how the set is arranged (by author name, or by subject—a conceptual surrogate), but a single physical surrogate can access only the document to which it refers. Conceptual surrogates support subject or term searches. Once the conceptual surrogates have been applied to a document, it can be retrieved using the controlled vocabulary of the surrogates. Conceptual surrogates also allow for many search results, rather than just the single reference of a physical surrogate. However, the flip side of the controlled vocabulary means that you must use the correct terminology, and employ it correctly, in order to return good results. Contextual surrogates, on the other hand, support full-text searches. Containing full (if brief) sentences, they provide the necessary material for natural-language searching. A search need only include a word used in the surrogate's narrative in order to retrieve the item.

True and complete information retrieval, however, is only possible when the different types of surrogates are employed together. For instance, a conceptual surrogate will be able to tell you that a certain document has been labeled with a term that interests you. However, it won't be able to tell you what the document actually says (contextual surrogate), or where it is or how many pages it has (physical surrogate). You can use a contextual surrogate to determine that an item is particularly important for pursuing a particular line of research, but it can't tell you what subject terminology should be used for it (conceptual surrogate), or whether your library has a copy (physical surrogate).

Surrogates, of course, do not apply only to the realm of documents held by a library or included in a database. For instance, a physical surrogate of Elmer Reiss might tell you that he was a dog (specifically a cockapoo) who lived in Seattle from 1971 to 1986, and that he had a white blaze on his nose, two white paws and a white patch on his chest. A conceptual surrogate of

Elmer might tell you that he was a family pet associated with bicycles, canoes, cross-country skiing and peanut butter. Elmer's contextual surrogate might say that he was a faithful and friendly soul who endured long hours with three girls in the back of a car, not to mention sharing his home and his people with two quirky cats, and that he was terribly confused when Dad moved out and took him to live only three blocks away.

Each type of surrogate is useful in its areas of strength. Each surrogate is needed for effective information retrieval. Each focuses on a different set of factors, and is used for a different purpose. Physical surrogates are the tools and materials of cataloging, which allows us to access items and know them individually. Conceptual surrogates are the means of indexing and devising subject headings, exercising the necessity of applying patterns of meaning to the world of knowledge. Contextual surrogates provide for abstracting, without which we would all be lost in a sea of publications without shore, seafloor, or hope of horizon. Surrogates act as our agents to organize and interpret information.