Tamara Marnell
SLIS S503
Lala Hajibayova
7/17/2010

Midterm

1) According to Blair, "effective document retrieval depends more than anything else on the linguistic representations used to describe documents". However, much of the data may be lost in the act of representing. Discuss the strengths and weaknesses of the three types of representation (formal, indexic and summary) for retrieval.

In an imaginary, perfect world, a patron could stroll into a library, pronounce what information they require, and a stack of ideal resources would appear. The stack would not omit any resources with pertinent information to offer, nor would it contain any extraneous resources of no value to the patron. Unfortunately, in this very real and imperfect world, this is an impossible task for even the most zealous of librarians. For one thing, patrons very rarely know precisely what information they seek. For another, the availability of information is limited by the authors who have taken the time to record it, the publishers who have taken the money to format it, and the powers that be who have given the library a budget to buy it. And most damningly, the necessity of representation for retrieval must chip away at the bulk of data held in a resource, potentially hiding its own value from the hapless OPAC-wielding patron.

Different types of representation have been fashioned to battle physical realities and human error, though none do so perfectly. Formal representation consists of attributes that describe the identity of a document: author, year published, title, etc. This kind of representation is closest to the presentation of raw data, and is the least vulnerable to indexer bias. Indexic representation, on the other hand, represents the conceptual or intellectual content of a document. Key words and subject headings are relevant examples. This kind of representation has the greatest power to aid or hinder the retrieval of documents. Finally, summary representation refers to natural language descriptors such as abstracts, reviews, or synopses. This kind of

representation is the most vulnerable to indexer bias, but can construct a more accurate representation of the content of a document than simple key words or subject headings.

The key strength of formal representation is its seeming objectivity. The MARC fields of a catalog record do not contain much room for debate; it is established fact that the publisher obtained copyright in a certain year, or printed a certain title on the cover. In the spectrum from data to knowledge proposed by Meadow *et al.* (2007), an expression of formal representation offers the closest to a datum, "a value of an attribute…[that] need not have meaning to everyone." However, precisely *because* a datum need not have meaning to everyone, formal representation is very inefficient for document retrieval. A patron who enters a library with a predetermined list of recommended titles or authors would find it useful, but the average individual browsing for exposure to new information would be completely lost if presented with a detached list of ISBNs.

Indexic representation adds the value of intellectual content that a browser needs. Subject headings and key words are theoretically optimal for controlled catalog searching, and have the pragmatic advantage of uniformity with the physical storage of documents. However, it is also the most dangerous kind of representation, because the injudicious application of a key word or subject heading has the power to shunt a document into the black hole of catalog invisibility. According to Buckland (1999) the very vocabulary used by indexers has a profound effect on search effectiveness. An analysis by Huber and Gillaspy (2000) of the vocabulary used in searching for articles on homosexuality in a medical database demonstrated this effect, and also brought to light the vulnerability of indexic representation to changing social norms and constantly evolving language. And Mai (2004) brought up yet another complication in indexing:

the difficulty of deciding whether to express an author's apparent intent, or to apply subject headings of optimal usefulness to users.

A compromise between the specific description of content and flexibility of search terms would appear to be summary representation. Abstracts, reviews, etc. allow greater strings of characters for users to potentially hit upon during a search. They also provide a more accurate report of a document's content, because an author can spell out his/her own purpose, results, and recommendations, instead of relying on the wisdom of librarians to apply a broadly understood subject heading. Unfortunately, these greater strings also open summary representation up to author or indexer bias. Ashworth (1973) described abstracting as "a fine art." Though to Ashworth, the "art" was to compact the essence of a document accurately and succinctly into a few lines of text, he acknowledges that it is basically a creative endeavor, and any creative endeavor is inevitably shaped by the biases of its creator.

These three forms of representation each have components that stand up to the messiness of reality and human error, but none are adequate on their own for the effective representation of documents in a library.

2) Discuss the difference between ordering and organizing.

The difference between "ordering" and "organization" is emphasized by Jacob (2004) in her criticism of the blurred lines between "classification" and "categorization" in discussions among information scientists. "A system for ordering," she begins, "provides access to resources by arranging them in some recognizable order." For example, a page of faculty descriptions on a departmental website listed by surnames is ordered alphabetically, or a series of bound journals arranged on the shelf by the date of issue is ordered chronologically. If a person is looking for a

paper published in that journal in 1963, he would be able to scan quickly past the books labeled "1950s" or back up after spotting the "1970s," and locate the correct volume between 1962 and 1964. On the other hand, "a system of organization…is a unified structure that establishes a network of relationships." Each volume in the hypothetical journal does not necessarily contain any intellectual content that establishes a closer relationship to its direct neighbors than to the volumes published many years earlier or later, so it is not "organized" within itself. A system of organization would apply, though, to that journal's close proximity to or shared classification with other journals of similar content in the larger context of the library's collection.

Two potential methods of organization are classification and categorization. Classification, Jacob asserts, "involves the orderly and systematic assignment of each entity to one and only one class within a system of mutually exclusive and non-overlapping classes." It fits the description of "organization" because the relationship of any entity in one class to a second entity in another can be determined based on the established features of the respective classes. Paradoxically, this reflects the classical view of *categorization*, in which each group is assumed to contain a unique set of qualities, and that any overarching categories inherit the qualities of its subgroups. The standard example of classification is the taxonomic web established by Linnaeus in his *Systema Naturae* (1735). Linnaeus' organization was also somewhat ordered, since his approach to the classification of angiosperms was simply to arrange them by the number of stamen and anthers on the flower (UCMP 2010).

Jacob's definition of categorization is a more compatible with the messy reality of mental representation, since the human mind more closely resembles a categorized cluster of overlapping Venn diagrams than a tidy classification tree. She explains, "Categorization is the process of dividing the world into groups of entities whose members are in some way similar to

each other." In essence, categorization is the "splitting and lumping" of objects and concepts described by Zerubavel (1991). It can only be considered a system of organization marginally, because it establishes an equality relationship between any entities *within* a category, but not necessarily *across* categories. As opposed to classification, categorization does not impose any strict boundaries on group belonging, so there are no set characteristics to establish relationships. In a traditional classification system, a paper on the effect of certain mouse genes on behavior must either belong to the Library of Congress *class* Biology—Genetics (QH426-470) or to Zoology—Animal Behavior (QL750-795). However, it is perfectly conceivable for the paper to be appropriate for both *categories*.

Categorization cannot be truly considered either a system of organization or ordering because it does not account for user browsing needs. For example, a student may enter a library searching for materials about mice for a term paper. The hierarchical structure of classes would allow him to navigate the sub-topics under "Zoology" to find documents on "Vertebrates," "Mammals," and finally "Mice," but in the indeterminate world of categories, gazing upon many books have the apparent similarity of being about animals would not direct him towards the best resources for his needs. Yet there is one angle from which categorization can be considered a 'stepping stone' to organization: it allows for post-coordinate indexing. Pre-coordinate, classified systems are inherently organized, but simply "splitting and lumping" documents by keywords can also allow a catalog searcher to create unique networks within a body of documents. An individual patron can establish a relationship between certain documents from the two categories of "genetics" and "animal behavior" by combining the two in an OPAC, or establish *exclusionary* relationships to other documents by inserting "NOT molecular" or something similar. In certain catalogs, like IUCat, he can also turn an indexer's original categorization into

an ordered system, by specifying that his search results be displayed from “New to Old” or “Author” instead of relevance.

Indexers should concentrate on organization over ordering in representation, as most people prefer to search for informative content rather than reading through long ordered lists of options. However, ordering is necessary for the physical storage of books on the shelves, since people should be able to locate resources by call number as painlessly as possible. Both have a place in the modern library.

3) Discuss how mental models (i.e. schemes, scripts and frames) of the environment serve as internal cognitive scaffoldings.

The concept of cognitive scaffolding was first introduced by the Russian psychologist Lev Vygotsky in the early 20th century, though he did not coin that term. Vygotsky proposed Zones of Proximal Development to model learning in children, defined as “the difference between the actual development level as determined by individual problem solving and the level of potential development as determined…under adult guidance or collaboration with more knowledgeable peers” (Gallagher 1999). This basic concept, that humans can augment their innate abilities through input from other people and tools in the environment, was later dubbed “cognitive scaffolding” by other developmental psychologists mid-century.

Mental models can serve as a kind of internal cognitive scaffolding. One type of mental model is the schema, described by Barlett (1932) as “an active organization of past reactions, or of past experiences.” Schemas are conglomerates of information amassed over a lifetime, ordered by experiential frequency to identify patterns, which are recalled when familiar stimuli appear. Another type is the script, a “frozen plan” (Schank & Kass 1988). Scripts are familiar to the public in the form of dialogue and acting directions for a movie, or an ordered series of

commands for a computer program. A third type of mental model is the frame, defined by Zerubavel (1991) as “the act of surrounding situations, acts, or objects with mental brackets that basically transform their meaning.” They are “experiential realms” that are largely socially constructed, and contain the rules for personal behavior and interpretation of others’ actions in situations like stage plays or games.

Each type of mental model—whether an individualized schema, a sequential script, or an abstract frame—is a way of preemptively understanding the environment. The purpose of constructing a mental model is to anticipate familiar structure in objects or actions, and therefore reduce the effort and time necessary to process stimuli. Bower (1996) acknowledged that the mental model of a stereotype can “function as thought-efficient starting points for understanding,” and Schank and Kass (1988) wrote that, “People process familiar situations much more easily than unfamiliar ones. They are very slow at processing novel situations, in which they lack specific expectations about what will come next.” The necessity of extra time to integrate new information into existing mental models is what we refer to colloquially as “surprise.” This element of surprise is often utilized by developmental psychologists, who can determine whether an infant comprehends physical phenomena based on the length of time they stare at presented situations (Goodwin 2010). Babies pay little attention to a ball slowing down as it fights gravity to roll up a slope. But they show surprise (i.e. stare at the set-up longer) if the ball is artificially *sped up* as it ascends that slope, because the phenomenon runs contrary to their mental constructs of normal ball behavior.

Constructs like this have obvious primal benefits to humans, such as being able to predict the movement of objects on hills (and move out of the way if, say, a boulder is perched precariously near the top of one). It also offers benefits in modern life. For example, though

social stereotyping can have negative side effects, it increases the efficiency of activities like electing political leaders. A person can safely assume that the representative of his favored party shares the bulk of his views without the need for detailed interrogation about individual issues. For a more specific example, if a college student is instructed in the creation of a mental model of, say, calculus, he can later draw on principles contained in that model to solve problems that he probably could not have figured out if someone had simply plopped $\int_1^2 3x^2 dx$ in front of him.

Such mental models of the laws of physics, political party lines, or mathematics reduce the cognitive load on an individual. This is the essence of cognitive scaffolding. According to Jacob (2010), the function of cognitive scaffolding is to "[minimize] the cognitive load…by constraining problem spaces, limiting the range of possible solutions, and providing criteria for selecting the most likely alternative." Mental models constrain problem spaces by establishing a bounded frame of reference to draw from. In the calculus example, a mental model would help the student eliminate all extraneous knowledge of mathematics that does not apply to the components of this particular integral problem. Models meet the second criteria, to limit the range of possible solutions, by storing previous experiences for comparison. The hypothetical student, through repeated exposure to similar problems in homework assignments and class demonstrations, should be able to recognize that the answer to this problem should be a positive integer, and that given those bounds, it should be within a certain range of smallish numbers. Finally, a model's organization of experiences into patterns provides proven strategies and criteria for selection of the best 'answer' or response to a situation. If the student has not memorized the exact anti-derivative of $3x^2$, he should recall the formula he was taught to derive it and subsequent steps necessary to solve the problem.

Mental models are more limited than other resources of cognitive scaffolding, because they are constructed entirely of individual experiences. External sources, such as teachers, technologies, and written aids (tables, formulas etc.) are of greater assistance when leaping from one Zone of Proximal Development to the next. However, mental models are still important components of the potential for human achievement.

References

Ashworth, (1973). The Information scientist, Vol. 7.

Bartlett, FC. (1932). Remembering: A study in experimental and social psychology. Cambridge, England: Cambridge University Press. p201.

Batty, D. (1989). "Thesaurus construction and maintenance: A survival kit." Journal of Academic Librarianship 12(6): 13-20.

Buckland, M. (1999). Vocabulary as a central concept in library and information science. Digital libraries: interdisciplinary concepts, challenges, and opportunities. Proceedings of the Third International Conference on Conceptions of Library and Information Science, Dubrovnik, Croatia, Lokve.

“Carl Linnaeus.” The University of California Museum of Paleontology (UCMP). Retrieved July 16, 2010 from http://www.ucmp.berkeley.edu/history/linnaeus.html.

Gallagher, C. (1999). Psychology History: Lev Semyonovich Vygotsky. Retrieved July 17, 2010 from http://www.muskingum.edu/~psych/psycweb/history/vygotsky.htm.

Goodwin, CJ. (2010) Research in psychology: Methods and design. Hoboken, NJ: Wiley. pp126-7.

Huber J & Gillaspy, M. (2000). An examination of the discourse of homosexuality as reflected in medical vocabularies, classificatory structures, and information resources. Dynamism and stability in knowledge organization: proceedings of the Sixth International ISKO Conference, Toronto, Canada, Ergon Verlag.

Jacob, E. (1991). Classification and categorization: Drawing the line. Advances in classification research. B. K. a. R. Fidel. Washington, DC, American Society for Information Science. 2.

Jacob, E. (2010). Overheads for L505—Augmentation. p3.

Mai, J. (2004). "Analysis in indexing: Document and domain center approaches." Journal of Information processing and management 41: 599-611.

Meadow, C.T., Boyce, B.R., & Kraft, D.H. (2007). Data, information, knowledge. Text information retrieval systems, 3rd ed. Amsterdam: Elsevier.

Norman, D. (1983). Some observations on mental models. Mental models. D. G. a. A. Stevens. Hillsdale, NJ: Lawrence Erlbaum. pp7-14.

Schank & Kass (1988). "Knowledge representation in people and machines." Meaning and mental representations. Bloomington: Indiana University Press. p.194.

Zerubavel, E. (1991). The fine line: Making distinctions in everyday life. Chicago: University of Chicago Press.