

Beyond Facets: Semantic Roots and Modifiers as Elements of a Conceptual Morphology

Dagobert Soergel

College of Library and Information Services

University of Maryland

College Park, MD 20742

This paper presents initial ideas on a conceptual morphology in which concepts such as Fermentation, Fermented, and Fermentable are represented as combinations of a semantic root, in the example Ferment, with a modifier, in the example process, state/condition, and susceptible to process, respectively. This makes it possible to generate a large number of concepts from a much smaller list of semantic roots and modifiers. It also allows for great flexibility in indexing and searching. The paper gives a preliminary scheme of modifiers and invites ideas from classification researchers, logicians, and linguists.

1. INTRODUCTION

Facet analysis identifies the semantic components of a concept and arranges the resulting elements into facets. For example

Hepatography = Liver & Imaging & X-rays

Hepatograph = Liver & Camera & X-rays

Hepatogram = Liver & Image & X-rays

The facets are **body part** (liver), **process** (imaging), **apparatus** (camera in the broad sense of any imaging device), **thing** (image, X-rays). This analysis results in great conceptual economy, since a given set of elemental concepts can be used to express a much larger number of combinations. Facet analysis also facilitates searching for broad concepts, such as *Camera (imaging device)* regardless of type, while still allowing for specific searches - e.g. for *X-ray camera* - by combination.

Three of the concepts used in the example, Imaging, Camera, and Image, while belonging to different facets, share the core meaning of *image*, which is broader than any of the three concepts. Some queries require this broad meaning, possibly combined with an imaging principle, such as X-rays. We can take facet analysis and conceptual economy a step further to support this type of broad search. We can express a concept such as Imaging by a combination of its semantic root - *Image* - with a general modifier, in the example process. Here are the results of such an analysis for the three sample concepts (semantic roots in italics):

Imaging = *Image* - process

Camera = *Image* - apparatus

Image = *Image* - resulting thing

Again, this analysis allows a broad search for the semantic root *Image* (combined with X-ray or ultrasound or whatever, as required), but also a specific search for Camera or, even more specifically, X-ray camera.

Figure 1 lists some other groups of concepts that share a semantic core.

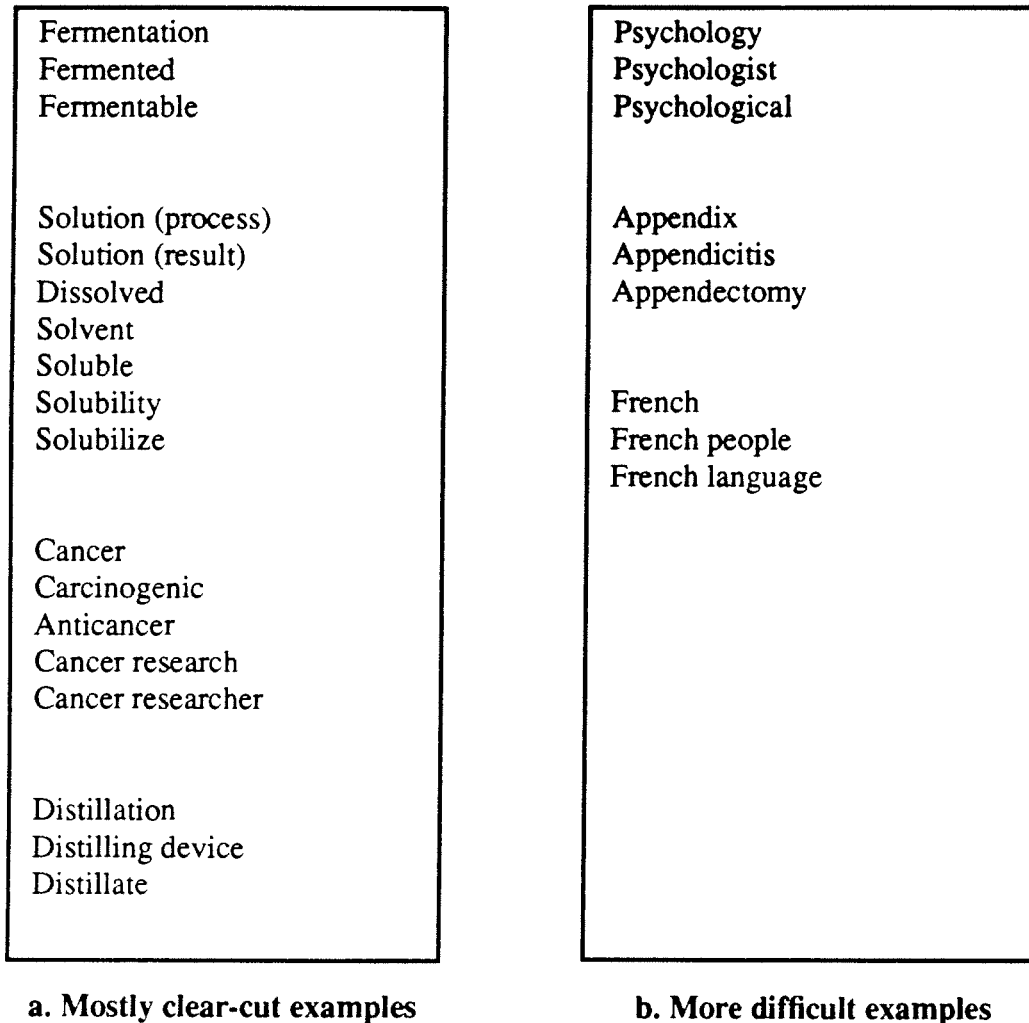


Figure 1. Groups of concepts that share a semantic root

The examples in Figure 1 suggest the modifiers already mentioned, **process**, **apparatus/device**, and **resulting thing**, as well as others, such as **susceptible to process or causing**. Other examples are modifiers that specify **comparison to a standard** (under, sub-; normal (default); over, super-) or **relative size** (micro-; normal size (default); macro-).

The principle of forming concepts by combining semantic roots with modifiers is, of course, borrowed from natural language, which uses roots and modifiers, such as **soluble**, to construct many words out of a much smaller number of elements. But natural language has many exceptions and irregularities, and some perfectly reasonable combinations (such as image-able) would sound quite unusual. I propose to apply this familiar principle in a kind of **conceptual morphology**. I envision a classification consisting of a well-structured skeleton of core concepts or semantic roots and a carefully worked-out system of modifiers. Many more concepts can be formed through combining these elements.

The task, then, is to develop a **logical system of modifiers that is expressive yet simple enough to be learned and used**. The paper provides the beginnings of such a system but is primarily a call for ideas from classification researchers, logicians, and linguists.

2. DEVELOPING A SCHEME OF MODIFIERS

One might assume that linguists would have worked out such a scheme in an analysis of modifiers used in natural languages, but consultation with a linguist produced no promising leads to work that would be applicable directly. So it seemed useful to make an independent start and then look at other work that is related and from which many useful suggestions can be drawn once the idea of a conceptual morphology has crystallized a little more. The task of developing a list of modifiers can be approached as a bottom-up thesaurus development task: Start with a sufficiently large number of concepts and perhaps arrange them into groups around core meanings. Then express each concept as a combination of semantic root and modifiers, making up new modifiers as needed. Often linguistic form will help in this process. Finally arrange the modifiers into a coherent schema and consolidate those that are very similar.

Analyzing a limited set of terms collected in conjunction with the Alcohol and Other Drugs Thesaurus resulted in the preliminary scheme presented in the appendix. The following examples illustrate the approach.

The first set of examples revolves around the semantic core *Ferment*.

Fermentation = *Ferment* - process
Fermenting

Fermented = *Ferment* - state/condition

Fermentable = *Ferment* - susceptible to process - state/condition

Thinking about fermentation, one can easily come up with other combinations for which there is no word in English:

Ferment - apparatus/device

Ferment - agent

Ferment - susceptible to process - process
(the process of making something fermentable)

The next set of examples revolves around the semantic root *solution*. Actually, this is a polysemous term or a homonym, depending on what commonality one perceives between problem solution and chemical solution. Most combinations are valid for either meaning, but there may be two English words or the English word may refer only to the problem-solving meaning or only to the chemical meaning.

Solving Dissolving	=	<i>Solution - process</i>
Solution	=	<i>Solution - resulting thing</i>
Solved Dissolved	=	<i>Solution - state/condition</i>
Solvent	=	<i>Solution - agent</i>
Solvable Soluble	=	<i>Solution - susceptible to - state/condition</i>
Solvability Solubility	=	<i>Solution - susceptible to - state scale</i>
(solvabilize) solubilize	=	<i>Solution - susceptible to - process</i>

Some of these examples illustrate the need for having two modifiers in a row. There may even be longer modifier chains. The concept to which the last modifier is applied is called the **base concept**. A base concept may be a semantic root or a combination involving one or more modifiers. It may even be necessary to admit base concepts which are constructed by combining several semantic roots with or without modifiers.

The modifiers are used to define concepts, not to make statements about reality. For example, *Cancer - agent causing* defines Carcinogen, which is a valid concept whether or not substances or other agents causing cancer actually exist. The statement:

Tar *isa* *Cancer - agent causing*,

on the other hand, uses the concept in making an assertion about reality.

The appendix gives a first sketch of modifiers to be used in conceptual morphology.

3. APPLICATIONS IN RETRIEVAL SYSTEMS AND LANGUAGE PROCESSING

As was mentioned in the introduction, the motivation for the development of conceptual morphology is two-fold: to allow for great flexibility in indexing and searching and to limit the number of elements needed in the index language. The first is achieved by being able to search broadly for semantic roots or specifically for semantic root modifier combinations. The second can be achieved by limiting the index language to the semantic roots and modifiers as the elemental building blocks from which the indexers and searchers can construct indexing concepts.

In some systems the specificity made possible by modifiers may not be needed. In that case, the index language can be limited to semantic roots. If some specific combinations are needed, they can be introduced as precombined descriptors.

An index language restricted to semantic roots and modifiers may put a burden on the indexers and searchers, who must understand the system and construct the appropriate combinations. This problem can be addressed by including many lead-in terms in the thesaurus or by introducing precombined descriptors. Both approaches are limited by what can be expressed in natural language. One might argue that a concept that cannot be expressed by a word or established phrase in the language is not important, but that is debatable.

Precombined descriptors not only facilitate indexing and searching but they also serve as examples on which indexers and searchers can pattern other combinations. On the other hand, precombined descriptors introduce a degree of inconsistency in that some combinations are already there, whereas others need to be created. Internally, precombined descriptors should always be resolved into their components for two reasons: (1) Entities indexed by a semantic root - modifier combination should always be retrievable by the semantic root alone, whether the combination is in the index language as a precombined descriptor or is constructed by the indexer. (2) Some indexers and searchers may find it easier to construct a combination than to use a precombined descriptor. This should make no difference to the system. This internal resolution into components also assures that all precombined descriptors are formed in the same systematic way, and that semantic relationships are thus made explicit.

The specificity and precision achievable by the approach described here should be particularly useful for the representation of "substantive" data (as opposed to bibliographic or other directional data). It may also be useful in dictionaries for natural language understanding.

4. OUTLOOK

The thinking about this system is still in an embryonic stage. The patterns of combinations and the modifiers listed in the appendix are those that most obviously suggest themselves from an analysis of a limited list of terms. The less obvious cases must be dealt with, and the somewhat ad hoc listing given below must be shaped into a coherent system that balances the need for expressiveness

against the need for simplicity. As was stated in the introduction, this paper is primarily a call to classification researchers, logicians, and linguists for ideas and references to related work which would serve as a basis for developing the schema further and apply it in some pilot systems.

APPENDIX

First sketch of modifiers to be used in conceptual morphology

In the following table, the left column gives a type of base concept and the right column gives modifiers that might be applied to that base concept. As noted in the text, the base concept can be a semantic root or a semantic root plus modifier. Usually it is clear to what base concept type a combination with a modifier belongs. For example,

Acid - addition

can become the process Acidification by further combining with the modifier **process** or the state/condition Acidified by further combining with the modifier **state/condition**.

Base concept	Modifier
Any concept	pertaining to that concept (adjective derived from the concept)
Any concept or topic	realm/discipline pertaining to the concept or topic (e.g., Heart - Cardiology)
Any concept, particularly Process State/condition Being/thing/object Substance	<p>agent causing or furthering (-genic) agent or process causing agent or process furthering</p> <p>agent killing or inhibiting (anti-) agent or process killing agent or process inhibiting</p> <p>Note: One could further specify these modifiers by inserting one of the following into the agent/process slot:</p> <ul style="list-style-type: none"> Process Thing/object <ul style="list-style-type: none"> Apparatus, device, machine Substance Being <ul style="list-style-type: none"> Organism Person <p>One might also want to look at the distinction between animate and inanimate agent and the related distinction between agent in a more narrow sense and instrument.</p> <p>susceptible to</p> <p>functionality`</p> <ul style="list-style-type: none"> good function (default) dysfunction (mal-, mis-) <p>Examples: liver - dysfunction (applied to thing; digestion - dysfunction (applied to process). A broad search could be (digestion OR digestive organs) -- dysfunction</p> <p>Note: Dysfunction could be subdivided into many specific modifiers for the type of dysfunction, particularly type of disease, but these then lose the intuitive character of a modifier.</p>

Base concept	Modifier
<p>State/condition Being/thing/object</p>	<p>process of development (-genesis, -poiesis) (Combinations can be further modified, see process/state/condition)</p> <p>process of reduction or dying reduction dying Note: <i>Killing</i> would be the modifier <i>agent</i> or <i>process killing</i> with the agent/process slot filled with <i>process</i>.</p>
<p>Process State/condition</p>	<p>process (If the base concept suggests a condition, this is the process that leads to the condition.)</p> <p>being/thing or state/condition (often the result of a process). being/thing state/condition state/condition as such (e.g., freedom) state scale state/condition as property of some being/thing (e.g., free) being/thing having that property</p> <p>degree of severity (related to Status of affirmation) not at all semi-fully</p>
<p>Being/thing/object seen in relation to a larger system</p> <p>Substance</p>	<p>addition</p> <p>removal, extraction</p> <p>replacement</p> <p>dysfunction/disease due to lack</p> <p>dysfunction/disease due to overexposure</p>

Base concept	Modifier
Any quantity	measurement of that quantity (may be further modified as Process/state/condition)
Real, area of endeavor field, discipline, occupation	<p>area/field/discipline/occupation as such, as a social organization, role of the area/field/discipline</p> <p>discipline</p> <p>profession, occupation</p> <p>person working in the area/field/occupation/</p> <p>pertaining to the area/field/discipline (special case of pertaining to)</p>
Space/time/figurative dimension (Concepts that can be used in two or three ways, such as before and under)	<p>space use</p> <p>time use</p> <p>figurative use</p>
<p>Geographical area/ethnic group/ group/language</p> <p>(There could be one list of semantic roots; each root may designate one or more of these, but see note.)</p>	<p>geographical area (e.g., France)</p> <p>ethnic group (e.g., French people)</p> <p>language (e.g., French language)</p> <p>Note: This one may be stretching the point. While the economy of having only one list replacing three, the relationship between France French people, and French language is more empirical than semantic. The hierarchy of languages in particular follows geography only loosely. So perhaps it would be better to maintain three different lists but add a fourth which would have concepts such as France, French people, French language to facilitate searching for all these aspects at once.</p>

Base concept	Modifier
<p>Any concept Very general modifiers</p>	<p>status of affirmation negated (non-, un-, in-) affirmed (default)</p> <p>presence vs. absence presence absence, without, -less</p> <p>comparison to standard under, sub- normal (default) over, super-</p> <p>relative size micro- normal size (default) macro-</p> <p>number of elements one, mono-, uni- two, dual, dyadic, bi- three, triadic, ternary, tri- four, quaternary, tetra- five, penta- six, hexa- etc. many, multiple, multi-, poly-</p> <p>temporal aspects before, pre- during (default), present after, post-</p>

User-Based Methods for Classification Development

Paul Solomon

School of Information and Library Science
100 Manning Hall, CB# 3360
University of North Carolina at Chapel Hill
Chapel Hill, NC 27590-3360

INTRODUCTION

Classification schemes fail too often because they are not grounded in the language and knowledge of users or in the task or situation of use. Yet methods for adding user and task dimensions to classification schemes are neither well established nor well tested. User and task grounding, however, are frequently vital, especially in situations where special language usage, knowledge, or experience either limit users' understanding or interpretation of a classification scheme or make a classification scheme inappropriate for particular users.

For instance, managers in a given organization in a particular industry will have ways of viewing their world that may not be obvious to outsiders. Even within a given organization, the differing perspectives of managers may lead to breakdowns (Winograd & Flores, 1986) in classification use or interpretation. The same is true at another extreme for school children whose restricted knowledge of the world leads to hazy and ad hoc conceptions that do not well support classification scheme development (Borgman, Chignell, & Valdez; 1989).

The question then is how to build an understanding of such factors as user language, knowledge, and task to ground classification schemes in their situation of use? This paper begins to address this question by focusing on two recent research projects. The first is recently completed research involving children using an online catalog (OPAC) in an elementary school library media center (Solomon, 1991). Among the motivations for this research were interests in understanding children's search language, the impacts of user search language on the success and failure of queries, the influence of the demands of the educational setting, the patterns of search term use across grade levels and stages of cognitive development, and the influence of use of Library of Congress Subject Headings (LCSH) as the classification scheme or index language to support information retrieval.

The second, a study of a resource allocation process in a Federal government agency, is entering the data analysis phase. This study was motivated by interests in reducing the time required to move through the resource allocation process and to allow a wider range of viewpoints to be considered in making allocation decisions. The initial idea was that these interests might possibly be served through cooperative work group software and decision aids.

THE CONUNDRUM OF CLASSIFICATION SYSTEMS

Both research efforts have the similarity of an existing classification scheme or index language that is employed by indexers, and in contrast, groups of users (children and managers) who either apply their own index terms or interpret those of the index language in ways that are different from the sense of the index language. The hope of index language designers is that the two points of view,

indexer and user, will be congruent and appropriate information retrieval will be facilitated. The conundrum is the potential for match between classification scheme or index language and user queries. Bates (1986) points to the problems of variety of search terms, the unpredictability of search requests, and the complexity of information access methods as fundamental issues in developing a design model for online catalogs or by extension in designing classification schemes to support user-based information retrieval.

USER-BASED METHODS

It follows that research regarding user-based classification schemes needs to investigate these problems of variety, unpredictability, and complexity for particular kinds of users or domains of information use. The question, then, becomes: How can a user perspective be added to our classification schemes? or Which methods lead to a sufficient understanding of user terminology, interests, task, and the like to allow classification development that meets user requirements?

Given this broad range of interests in understanding variety of term use, unpredictability of queries, and complexity of information retrieval mechanisms, methods that allow classification researchers to capitalize on the rich natural environments of classification use for information retrieval are indicated. That is, methods that simplify the environment of classification use for information retrieval also mitigate variety, unpredictability, and complexity and leave the researcher with an incomplete view of the demands placed on the classification scheme.

Fortunately, there is a set of naturalistic or qualitative methods (e.g., Goetz & LeCompte, 1984; Merriam, 1988) that allows researchers to mine the natural information environment and to begin to understand the heterogeneity of user behavior within that context. By understanding this heterogeneity of user behavior the researcher can begin to discern patterns of behavior that, for instance, lead to user-based classification concepts, lead-in vocabulary to support thesaurus development, or an understanding of situational and conditional variations in term use necessary to add intelligence to a computer-based classification scheme and the information systems that employ it.

Typically the naturalistic researcher's tool kit consists of such methods as observation, interviewing, and document analysis. The use of these methods is interactive: Observation leads to the need to ask questions or to study written procedures or policies. An interview or analysis of written material may suggest something that needs to be emphasized in observations. For the purposes of classification research the focus might be a transaction at an online catalog, a query of a database, or an iteration of a resource allocation process. Surely major interests for the classification researcher in observation and questioning will be how a user employs any classification interpretation tools (e.g., a manual, list of subject terms, a graphic) and what terminology or category a user selects. However, the challenge for the researcher is to go beyond the readily visible surface action and delve deeper. One approach for accomplishing this is to ask the user to think-aloud while performing the task (Ericsson & Simon, 1984; Fidel, 1984). Another is questioning the user about, for instance, term choice, classification interpretations, or their information problem.

It turns out that by performing such investigations the variety in term use begins to become defined and supplemented by an understanding of similarity in term use. Also, the unpredictable of encounters with an information system begins to become predictable and the complexity of information retrieval mechanisms begins to become understood.

CHILDREN, AN OPAC, AND LCSH

This research explored the intentions, plans, opening moves, strategies, terminology, and successes and failures of children using an Online Public Access Catalog (OPAC) in a School Library Media Center. The research site, Bonnie Brae Elementary in Fairfax County, Virginia, was in its second year of operation with an online catalog and no card catalog. Children spread through grades one through six were observed to use the OPAC successfully in about 66 per cent of the observed OPAC transactions.

From the point of view of classification research, the emphasis of the research was on understanding the interplay between the search terms generated by children and those LCSH available in the collection. The research employed the range of qualitative methods already described. Interest in the role of classification in the total information retrieval process at the OPAC led to detailed attention to children's search terminology. That is, there was a need to know more about children's search terms than just what they were. There was a need to know something about the impetus for the search (e.g., personal interest, assignment), the decision process used in arriving at the term(s) entered at the OPAC (e.g., "It just came into my head," "That's the word my teacher used," "I thought about using clothes, but it doesn't usually like my words. So, I tried to think of other words like dress and came up with costumes."), and the thinking that occurred when an initial term was unsuccessful (e.g., "Well, I tried weapons and got one about space weapons. Then I tried guns and rifles.").

The research process of interaction between data collection through various means (observation, questioning, children's think-alouds) and data analysis (development and expansion of field notes, identification of patterns of behavior, generation of particular questions for follow-up in further iterations of the research process) allowed the researcher to begin to understand over time the nature of variety, unpredictability, and complexity in children's use of the OPAC in the Bonnie Brae Media Center. For instance, the problems of variety and unpredictability in terminology became much more manageable when it was understood that children in the lower grades had certain common interests in animals and sports that resulted in frequent application of the same terms. Even for children in the upper grades who were much less likely to use overlapping terms, the research process led to an understanding of the interests of these children along with the demands of the curriculum and, thus, enhanced the predictability of search requests. Classification researchers might use knowledge of children's interests and the terms they employed to suggest classification displays or tools that would help children access the information about book availability in an OPAC as supplements to a more general classification like LCSH.

In short, there were many findings that demonstrate possible classification system supports for the child users of OPACs in schools. For instance, many children, especially in grades one, two and three, showed a propensity to use simple, concrete terms (e.g., dogs, cats, dinosaurs) that directly mapped to LCSH and, consequently, resulted in search success. The message for the classification

researcher or designer here is that design attention can better be focused elsewhere, but that classification displays or tools leading to other related terms for these interests may have broad appeal. Other children, especially in grades four through six, showed a propensity to use complex, abstract, idiosyncratic terms (e.g., scary, horror, freedom, slavery) that did not map to LCSH and, thus, resulted in search failure. This finding suggests areas where classification researchers/designers can productively focus attention to overcome inadequacies of LCSH with respect to children as users.

Patterns of interest among children and patterns of classroom assignments, when coupled with an understanding of the associated breakdowns in information retrieval that children incur, also indicated areas where classification maps or other user supports would be helpful. That is, the fact and circumstance of the breakdown (e.g., that children used weapons and the LCSH was arms) suggests possible classification tools (e.g., a mapping between weapons and arms). The variety of observed breakdowns suggest many possible improvements. LCSH may simply not cover an area of interest to children (e.g., motocross). LCSH may employ terms that children do not normally use (e.g., hygiene for health). Children may need to employ a variety of terms to fully access information in a subject area. This last possibility proved particularly troublesome to children in that they were often unable to come up with the LCSH.

Researchers in the fields of special languages and terminology have reported that some concepts are consistently applied through single terms with no alternate terms or synonyms (terminological consistency); other concepts suffer from the existence of many alternate words or synonyms (terminological diversity) (Bhattacharyya, 1974; Svenonius, 1986). A particular distinction recognized by these authors as well as Bonzi (1984) is that words having no synonyms are largely concrete in nature, while words with many synonyms tend to be abstract. To the possibility that a term may have synonyms, another possibility must be added: a term may have multiple meanings, as does bat (animal and baseball). The distinction of single versus multiple meanings is labeled semantic consistency/diversity, or polysemy.

One example of how these issues of terminological and semantic consistency/diversity influence children's success at an OPAC comes from the domain of airplanes where children tended to use the term planes initially. This use of the term planes led them to only a small subset of items containing planes as part of a multi-word term (e.g., jet planes, fighter planes). Children's use of other common entry terms (e.g., flight, air) led to some relevant items, but also to items on birds and fresh air. The point is that an understanding of the problems that users face in actually employing an information system that relies on a classification system like LCSH allows the development of classification interpretation tools for an area of interest. Availability of these tools, in turn, can lead users to a wider range of relevant information than they would otherwise be able to access.

Another example comes from the distribution of subject headings for Black History. Black history month is celebrated each January and virtually every student in the school is given some related assignment usually involving the Media Center and the OPAC. In using the OPAC children tended to enter the term Black or Black Americans, which got them no where. They needed to know that the subject headings in use in the collection were Blacks and Afro-Americans. It may have also been helpful for children to know that the names of individual Black Americans (e.g., Martin

Luther King, Bill Cosby) were included as subject headings and that those persons were usually assigned an occupational subject heading (e.g., football players, comedians, clergy). Again, knowledge of information needs and the terms that children naturally employ when dealing with an information need give researchers and designers insights that can be used in classification schemes and interpretation tools.

The research demonstrated that there are a range of methods that may be used to add user and task perspectives to a classification scheme. Observation techniques point to the context of successes and breakdowns or the situational patterns that lead to success and breakdown. Breakdowns and their impacts are particularly informative in specifying the conditional aspects of a classification or the conditions when alternate terms may be required. Yet, the context of success and failure gives only part of the picture. Questioning techniques including think-aloud protocols try to pinpoint the patterns of behavior that lead to breakdowns. The interests of children in first grade led them to use concrete terms (e.g., insects, karate, snakes) that directly map to LCSH; the information needs of sixth graders led them to use more complex and abstract terms (e.g., slavery, freedom, ozone depletion) that frequently do not map to LCSH. Documentary sources including the terminology employed in a search provide a baseline of information to discern patterns of behavior in connection with the other research methods employed.

Use of these methods and associated analytical procedures led to identification of areas where classification scheme amelioration would be particularly beneficial. For instance, areas where a child needs a variety of terms to fully access a subject area (e.g., the domains of games, the military, and martial arts) were identified. The research approach also led to the indication of approaches to classification improvement that would have meaning for the particular class of users of focal interest in the research: children. For instance, graphical displays of term relations seem most appropriate for conveying the structure of some subject areas (e.g., airplanes) and faceted lists for others (e.g., black history).

MANAGERS AND A RESOURCE ALLOCATION CLASSIFICATION

Many organizations and units of organizations go through a periodic process of making decisions regarding the allocation of funds and people to particular activities. In the organization that was the site of the research reported here, a unit of the National Park Service, the annual process for the allocation of resources to natural resource conservation projects at state and local levels has been a time-consuming, almost paralyzing procedure that eventually results in satisfactory project selections. However, the process so debilitated the staff that follow-up activities such as project evaluation and reporting were avoided.

The primary purpose of the qualitative study of the 1991 version of this resource allocation process, then, was to discern what aspects of the process led to this near paralysis including the behaviors of participants, the kinds of conflicts that arose in making allocation decisions, and the specification of elements of the group decision process that were easy to make and those that were difficult. Ultimately, a fundamental goal of the research was to possibly specify computer supported work group solutions to whatever sources of breakdown were identified.

The research process consisted of collection of memoranda, notes, and other written documentation; observation of group decision meetings; and interviews with participants in the process. Field notes were prepared for each observation and interview session. Transcripts of tape recordings of the group meetings and interviews were also made. Interim analysis of the data in the form of expansion of field notes and transcripts was performed following the development of field notes for observation and interview sessions. Research memoranda were written periodically to summarize the research to date and included both descriptions of patterns of behavior and questions that needed to be investigated further through observations, interviews and document analysis. Final analysis of the data is in progress.

It became apparent during the interim analyses that one source of the time consuming conflict in allocation decisions was the simple classification system that had been in use for many years. This classification system directly lists seven kinds of projects: river corridor, trail corridor, state river assessment, regional trails plan, metropolitan area trails plan, training/consultations, and the inevitable other. Different managers had distinct interests in reviewing projects that reflected their management responsibilities. One manager focused on trails; another had responsibilities for rivers. Still another was concerned with a variety of balancing considerations: for example, geographic balance, funding requirements, time requirements.

The application of this classification scheme, thus, inevitably brought about conflict in the form of discussion about who had oversight responsibility for the project especially when a project transcended one resource type into another and perhaps extended into broader uncovered types (e.g., landscapes or greenways). The point is that the classification scheme, while seemingly clear-cut and simple to use, did not well support the kinds of priority decisions that these managers needed to make.

Preliminary analysis of the user-based data led to identification of a variety of management concerns, resource allocation priorities, and political realities as important decision factors. Factors initially identified included: management resource requirements--likely demands to be placed on management for oversight; scope of effort--project has effect at local, regional (substate), state, multi-state, national, or international levels; resource requirements--length of time and human resources required to completion; and resource characteristics--river, trail, greenway, etc.

In summary, systematic study of the resource allocation process identified a number of problems with the process, including the incongruence of the simple classification system with the decision making factors used by the collection of individual managers. The user-based research approach employed helped to identify the concerns of individual managers and to understand how individual concerns hindered the collective decision making process from quickly reaching a conclusion. The concerns of these managers, in turn, enable the systems designer to invent a classification scheme that promotes the decision making process.

TAILORING CLASSIFICATION SYSTEMS TO PROMOTE USER SUCCESS

The implication of these examples of user-based classification research seems obvious: Ultimately, any classification system or tool needs to be evaluated in situ. Whether the classification system

or tool results from intellectual, linguistic, or algorithmic analysis of some domain of interest, the demands put on the classification scheme by real people having real information needs will at times diverge from the expectations of designers. The case study of children using an OPAC employing LCSH discussed above illustrates this point.

Some parameters of success and failure have been identified: Children's queries usually succeeded when simple, basic-level terms were employed (e.g., dogs, cats, horses) and usually broke down when terms representing complex, abstract concepts were employed (e.g., scary, feelings, thriller). Terminologically diverse concepts frequently led to breakdown because of the wide range of terms that can be employed to refer to a concept: bike, bicycle, BMX) (Bonzi, 1984). The special meanings that underlay some terms (e.g., dinosaurs were viewed as monsters by children in the early elementary grades and not animals) led to confusion when children tried to exploit hierarchical relations (child: dinosaurs→monsters; adult: dinosaurs→animals). The information needs that provided the basis for children's queries, including those representing the interests of children themselves as well as those representing the educational environment and its curriculum, were sometimes well covered within LCSH and sometimes not. This situation promoted success when subject headings were available and breakdown when they were not.

The point is that knowledge about user behavior in relation to a classification system like LCSH gives classification researchers a variety of insights that can be used to develop classification tools and schemes that are tailored to user behavior and information needs. The availability of tailored classification schemes and tools seems especially important when dealing with classification systems like LCSH that are so well established that they may be impossible to change. Even if change were feasible, the likelihood of successfully developing a general purpose classification system, which is used in so many different situations for so many different purposes, seems low. Rather special tailored classification tools and schemes can be developed to fix the gaps and meet the particular problems and needs of users, if these gaps, problems, and needs are known.

In contrast, the understanding of the classification needs of managers provided by the case study of the resource allocation process gives the classification researcher the opportunity to invent a new classification scheme or remodel the old. The experience of the case study suggests that a multiple dimensional classification is needed to satisfy the diverse interests and information needs of the users involved. By faceting the interests and concerns of the managers involved, the unidimensional classification becomes less fuzzy and highlights key concerns in the resource allocations process: management requirements, scope of effort, resource requirements, and resource characteristics.

The two examples using qualitative methods presented above show that such methods have potential for adding a user dimension to classification research. Both cases made use of existing classification systems as a benchline for understanding user success and breakdown and associated patterns of behavior. Frequently, a research setting will have some classification scheme already in use. In other situations the classification researcher will provide a classification system or set of tools to be tested in relation to some information setting and system. Less frequently, there will be no starting point in terms of classification and the research effort will focus on classification development. In each of these cases qualitative methods provide the researcher with the flexibility and control to describe, evaluate, and understand people's classification needs and behavior.

However, those employing qualitative methods for classification research need to recognize that these methods put great demands of time and intellectual resources on the researcher. They require considerable planning and organization throughout the research process as the researcher considers and tests alternate descriptions and explanations for user behavior. Just the process of developing field notes is often exhausting. Employing those notes for analysis and reporting purposes may be even more burdensome depending on the extent of computer support available (Tesch, 1990). Yet, for those classification researchers interested in promoting the usability of their classification approaches there are few, if any, viable alternatives to testing the products of their research with real people in real settings of use.

REFERENCES

- Bates, M.J. (1986). Subject access in online catalogs: A design model. Journal of the American Society for Information Science, 37(6), 357-376.
- Bhattacharyya, K. (1974). The effectiveness of natural language in science indexing and retrieval. Journal of Documentation, 30(3), 235-254.
- Bonzi, S. (1984). Terminological consistency in abstract and concrete disciplines. Journal of Documentation, 40(4), 247-263.
- Borgman, C.L., Chignell, M.H., & Valdez, F. (1989). Designing an information retrieval interface based on children's categorization of knowledge: A pilot study. Proceedings of the 52nd Annual Meeting of the American Society for Information Science, 26, 81-95.
- Ericsson, K.A., & Simon, H.A. (1984). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT.
- Fidel, R. (1984). The case-study method: A case-study. Library and Information Science Research, 6(3), 273-288.
- Goetz, J.P., & LeCompte, M.D. (1984). Ethnography and qualitative design in educational research. Orlando, FL: Academic Press.
- Merriam, S.B. (1988). Case study research in education: A qualitative approach. San Francisco, CA: Jossey-Bass.
- Solomon, Paul (1991). Information Systems for Children: Explorations in Information Access and Interface Usability For an Online Catalog in an Elementary School Library [Dissertation]. College Park, MD: University of Maryland.
- Svenonius, E. (1986). Unanswered questions in the design of controlled vocabularies. Journal of the American Society for Information Science, 37(5), 331-340.
- Tesch, R. (1990). Qualitative research: Analysis types and software tools. Philadelphia, PA: Falmer Press.
- Winograd, T., & Flores, F. (1986). Understanding computers and cognition: A new foundation for design. Norwood, NJ: Ablex.