II. Assignments

Assignment 1: Abstracting

1) 1st article: indicative abstract

Date: 1/14/02

David, C., et al. (1995). Indexing as problem solving: a cognitive approach to consistency. In T. Kinney (Ed.), *Forging New Partnerships in Information: Converging Technologies: Proceedings of the 58th ASIS Annual Meeting* (p. 49-55). Medford, NJ: Information Today for the American Society for Information Science.

The researchers apply principles of cognitive psychology and problem solving theory to indexing of documents, with the aim of shedding light on issues of interindexer consistency. Previous research has found low rates of consistency between different indexers approaching the same document. This article reports on results from four expert indexers analyzing one article with one thesaurus (partial results from a larger study). The indexers were videotaped while working, and were prompted for three types of verbal report: concurrent ("think aloud"), retrospective (follow-up interview), and peer evaluation (comments on the other indexers' choices). The goal of the researchers is to model the intellectual process of problem solving for the task of indexing, and thus to explain variation in choice of indexing terms.

2) 1st article: informative abstract

Date: 1/15/02

David, C., et al. (1995). Indexing as problem solving: a cognitive approach to consistency. In T. Kinney (Ed.), *Forging New Partnerships in Information: Converging Technologies: Proceedings of the 58th ASIS Annual Meeting* (p. 49-55). Medford, NJ: Information Today for the American Society for Information Science.

Using problem solving theory from cognitive psychology, the indexing task is determined to be an ill-defined problem, since the nature of the ultimate goal is defined by the individual indexer. Variation occurs even among experts due to personalized standards and preferences, which lead to divergent decisions regarding appropriate terms. A relatively high rate of consistency between expert indexers was expected, due to similar background and skill-levels. However, the rate of interindexer consistency among sampled experts came to only 45% overall, with 66% agreement for the primary descriptors. It was found that the experts followed differing criteria for the selection of terms—that they perceived the priorities of the task and the needs of their

audience differently. These initial results validate the application of problem solving theory to explain the intellectual processes of indexing.

3) 2nd article: indicative abstract

Date: 1/15/02

Buckland, M. (1991). Information as thing. *Journal of the American Society for Information Science*, *42*, 351-360.

The author discusses and clarifies different definitions surrounding the idea of "information" and its subcategories. {Note: technically, I think I could stop here.} He defends the technically incorrect usage of information-as-thing (as an object or resource), since information systems can deal directly with information only in this sense. He relates information-as-thing with "evidence", and goes on to delineate different types of information and the problems associated with the term "document". He includes "events" as a kind of information, and stresses that information is situational. He notes that, given the definitions he outlines here, anything could potentially be informative, and thus be information.

4) 2nd article: informative abstract

Date: 1/15/02

Buckland, M. (1991). Information as thing. *Journal of the American Society for Information Science*, *42*, 351-360.

"Information-as-thing", where information is considered an object or resource to be retrieved, discovered, or used, has risen in importance as the common definition of "information", and can now be seen as a valuable means of organizing information study. Traditionalist scholars have rejected this usage, however, since it does not reflect the original meaning of the word. The significance of information-as-thing is defended, since information systems can deal directly with information only in this sense. Information-as-knowledge and information-as-process (the two traditional definitions) are intangible elements of information; information-as-thing, and a fourth element, information processing, are tangible. Information-as-thing can be equated with the idea of "evidence". Four types of information resources are distinguished: data, documents, objects and events; controversies surrounding the term "document" are noted. It is observed that, if information is evidence and if evidence is whatever one might learn from, anything could potentially be or become information. Further, it is stressed that information is situational, and consensual. The nature of copies and of representation is discussed. Representations of knowledge are held to form a subset of information-as-thing. "Information-as-process", with "information-as-knowledge" as its subset,

remains important as a focus for information study. Information-as-thing is a valid concept that will not supplant prior understandings of "information", but can be used to bring shape to the discipline.

5) 3rd article: indicative abstract

Date: 1/15/02

Randi, J. (1996). Investigating miracles, Italian-style. *Scientific American* (February 1996), 136.

Randi decries the trend in modern Italian popular culture celebrating miraculous phenomena involving weeping or bleeding statues of the Madonna. He describes laboratory experiments by Luigi Garlaschelli and colleagues that offer physical explanations for these and other events involving blood, both contemporary and historical, which have been hailed as miraculous.

6) 3rd article: informative abstract

Date: 1/15/02

Randi, J. (1996). Investigating miracles, Italian-style. *Scientific American* (February 1996), 136.

The popular phenomenon in modern Italy of statues of the Madonna that appear to weep or bleed is a disheartening throwback to medieval superstition. The phenomenon is described, and scientific explanations as demonstrated by Luigi Garlaschelli and colleagues are offered. A small hole can be drilled in the top of a plaster or ceramic statue; the statue can then be filled with liquid and drained. Since plaster and ceramic are porous, the material retains some liquid that can then flow from the eyes of the statue, causing it to appear to be crying or bleeding. Since owners of these miraculous statues refuse to allow either the statues or the tears/blood to be tested, Garlaschelli's theories cannot be proven. Historical instances of miraculous blood appearing on food items are explained by the agency of a bright red microorganism, *Serrratia marcescens*, which thrives on starchy foods in warm, moist environments. Garlaschelli and a colleague have also experimentally debunked another purported miracle involving blood, but once again the sacred subject matter allows the church to oppose direct proof.

Assignment 2: Information

Date: 1/30/02

For the broadest definition, I follow Michael Buckland's idea of information as "evidence" to claim that information is how we learn things—that through which we acquire or receive material by which we say we understand, we believe, we know. Sometimes we seek it for specific purposes; sometimes we come across something that relates to a subject in which we were already interested; sometimes we meet something we hadn't had in mind that stimulates new interest or a new project. We can obtain information through formal sources intended to contribute knowledge; we can receive information through interaction with other individuals, whether verbal or non-verbal; we can be informed through our senses, by observation of the environment. What the information means depends less on intrinsic content, and more on who is responding to it and what their interests are. Information is in the most general sense communicated, coming from its source to a recipient; but it is not always received or acquired through a deliberate act of direct communication.

For information professionals and for many people who rely on "information" to meet their professional and personal goals, the definition becomes more specific and more practical. "Information" comes into focus in its role as a commodity or a resource, again corroborating Buckland in his pursuit of "information-as-thing". Information in this context is the meaning-bearing material that must be processed, stored, retrieved, and often processed again—the stuff we can analyze and code into databases and provide access to, the stuff we search for, look up and refer to. It is data; it is facts we want to know or want others to know. In this sense, information is a tool, an asset; it is currency and even wealth. We seek information to answer a question, to aid a decision, to shape a plan, to produce a report. Traditionally such information has been alphanumeric, conveyed in text and numbers; but these days we encode and access sound and both still and moving images as well (of course, so far we still usually need some kind of alphanumeric "handle" to get at these other media).

Information is not just for humans, although we usually think of it as human artifact and human tool. Animals and plants and even stars and magnetic fields also generate and receive "messages" through signals that must overcome systemic or environmental noise, to borrow Shannon and Weaver's model. For our purposes as students and future representatives of library and information science, however, the human context is more than sufficient. And the human context can be rich and complex. Seely Brown and Duguid have rightly promoted the idea of "the social life of information", emphasizing that information is not autochthonous or self-generating, but exists in the context of human activity. People both produce and use the items we typically think of as information, and they tend to interpret those items in terms of their personal needs as well as their interactions and relationships with other people—in a personal and a social context.

Stepping outside and noticing the temperature gives you information. Looking at the gas gauge on your car gives you information. The feel of a fabric, the smell coming from the mystery container in your refrigerator, the sound of your dog's toenails coming toward you down the hall—all these things provide evidence and allow one to draw conclusions or form opinions, to know something one didn't know prior to the encounter. Of course, so do dictionaries and databases and websites and quarterly reports, not to mention conversations or scribbled post-it notes or the tone in your voice when you're annoyed that reminds you so frighteningly of your mother. If Buckland is right, and everything can be informative (and it's hard to tell him he's wrong), then we will never be finished trying to define information.

Assignment 3: Representation

In his book *Things that make us smart*, Norman used the example of computation with Arabic numerals as compared with Roman numerals to discuss issues of representation. What are the advantages and disadvantages of different systems? Why is one chosen or used instead of another? This brief essay will examine the parallel system of the Roman alphabet we use in the European-derived world today, as compared with Chinese pictographs as a system of writing.

The Roman alphabet was based on the Greek system, which was in turn based on Phoenician writing, which for its part developed from Sumerian symbols used to record numbers and events. While these alphabets look different, they are based on the same principles: letters approximately and flexibly represent sounds, and are combined using rules of spelling to represent words in sentences. Chinese writing, by contrast, uses symbols to represent entire words. Every word has its own symbol, although symbols can be made up of component segments that carry their own meanings. What are the advantages of or problems with either system?

On the face of it, an alphabet is easier to learn to read and write than a pictograph system like traditional Chinese writing. The Roman alphabet has twenty-six letters, with which all words are formed. Learn those twenty-six letters, and you know the system. For Chinese, one must learn a new symbol for each word. With letters, the visual representation has an association with the sound of the language, providing a guide to how the word is pronounced; pictographs relate not to sounds but to concepts. To read Chinese aloud correctly, one must know not only what the word is, but also what its sounds are. Phonics won't help you learn to read Chinese.

But is written English really so very different? It doesn't take a student of English as a foreign language to tell you that the rules of spelling and pronunciation are often arbitrary at best. The principle that ontogeny recapitulates phylogeny applies to the Roman alphabet as well as to embryos: the history of the language has left its mark on our spellings. Obsolete pronunciations and centuries-gone linguistic influences (such as Norman French) are fossilized in English spellings, making many words either a history lesson or simply a frustrating example of the human propensity for illogical commitments. Not only must one learn the rules of spelling, one has to learn and recognize all the exceptions. George Bernard Shaw was not just being facetious when he suggested "ghoti" as a proper spelling in English for the word that indicates an aquatic animal with fins and gills.

Initiatives for written English reform to bring spellings more into line with pronunciation have periodically been proposed, but beyond Webster's assertion of more "sensible" American spellings (such as "color" vs. "colour", "theater" vs. "theatre") and informal usages such as "enuf," "cuz," and "altho," they have never taken real hold. While letters roughly represent sounds, the sounds they represent change over time and from place to place. The English

that Shakespeare spoke sounded very little like what we now recognize as Shakespearean English. I used to play a little game with a friend from Australia and a friend from Rochester, New York, where we would each say the word "bomb". In my flat Pacific Northwest accent, it came out approximately "bahm"; my Rochester friend had a sharp nasal o that came out something like "bam" to my ear; and our Australian friend said "boamb", with a long o and pronouncing the final b.

For many of us, as experienced readers we negotiate English words in print not so much by piecing together the letters as by recognizing the "look" of a word as whole—the same way a person literate in Chinese reads the pictographs. In order to read either language, one has to be able to associate words with concepts or things—to translate the written markings into linguistic sound and meaning. Why, then, do we have an alphabet instead of a pictograph system?

The simplicity of the system has already been mentioned. In part we use the Roman alphabet simply by an accident of history: the Roman empire and the Latin language left an enduring mark on European culture. The Romans got their alphabet from the Greeks, who got theirs from the Phoenicians, whose own system developed from early Sumerian symbols. And it's turtles (or history) all the way down. The Chinese empire was at least as sophisticated as the Roman one, however, and in many ways the writing system is far more sophisticated. Its elite nature—years of education are required to master it completely—was in keeping with the Confucian social structure that supported the empire.

But then we invented the printing press, and movable type, and thence the typewriter and the keyboard. An alphabet is far easier to manipulate in rapid mass production of this kind than a pictograph system. It is not surprising that in the twentieth century simplified versions of Chinese characters have been developed, and also a regularized system for transliteration in the Roman alphabet, called pinyin (of course, it helps to have a totalitarian government if you want to implement such major reforms).

For something as ancient as writing, we use a representational system that we have inherited, that comes with the territory, for better or worse. This exercise could have been performed with younger, more newly invented systems, such as Library of Congress call letters, American traffic signs, or the system of icons used in the Windows computing environment. I chose written language in part to echo Norman's work, and in part due to my own fascination with language. Written language is the lifeblood of academic work, and thus an appropriate subject for academic examination.

Assignment 4: Grocery Store Script

Date: 2/17/02

I. Script

1. Arrive at store and proceed to entrance.

2. Choose basket or cart for carrying groceries while shopping.
    a. basket for intended quick trip with fewer items or many small items
    b. cart for more extended trip including larger, heavier items

3. Proceed through store selecting items.
    a. If pursuing only specific items, target those areas of the store only.
        1) Proceed following mental or written list of needed items.
        2) If using coupons, seek items according to coupons held.
    b. If making a general trip, go through each aisle or section, seeking useful items (omit aisles containing items not applicable to lifestyle, e.g., baby or pet supplies, liquor section, bakery, etc.).
        1) Coupons may be combined with this step as well.
        2) Scan for sale items to trigger purchase according to value of discount and desirability of item.
    c. Strategies a. and b. may be combined for a hybrid trip.
        1) specific hybrid: select general items only when noticed near specific items desired. Selection triggered by recognition of additional needed item, or good bargain for useful item, in the vicinity of a specific item already on mental or written list or included on coupon(s).
        2) general hybrid: a general trip may be intended, but specific items must be acquired during the general search of the store. A mental or written list or coupon(s) may be followed for the specific items sought during a general hybrid shopping trip.

4. Check out at cashier.
    a. Offer any coupons or store discount cards (bar-coded).
    b. Specify paper or plastic bags.
        1) If you have a particular preference, and the bagger does not ask, volunteer your preference.
        2) If no bagger is present, or the cashier does not begin to bag groceries for you, bag items yourself.
    c. Make payment by cash, check, debit card or credit card.
    d. Accept receipt.

5. Leave store with groceries.
    a. If shopping with basket, carry your bags with you. Basket remains in store.
    b. If shopping with cart, take cart with groceries to your mode of transport (normally your own car, but may be a friend's car, taxi, or bus). Place grocery bags in vehicle, and return cart to designated area in parking lot.

II. Analysis

The basic outline here should apply to any grocery store type experience, including specialty shops such as Sahara Mart or Saraga Oriental Market. I had the E. 3rd St. Bloomingfoods in mind, however. Instantiations of this schema will vary according to the individual's accustomed shopping strategies, the nature of a particular shopping trip, and sometimes with the organization of a particular store. For instance, Kroger's now features a Self Check-out option for those carrying about 15 items or less. This changes the check-out process slightly, since one is not interacting with a cashier or a bagger but with a machine (and an attendant, if anything goes wrong).

From my own experience, I do most of my grocery shopping these days at three different places: Sam's Club, Bloomingfoods, and Kroger's. I will also make quick trips to Osco to get milk and medicines sometimes (my husband has allergies, so nose spray is always on the list), since they have low prices for milk and there is an Osco's on my way home. I shop very differently at my three main stores, however, in part because the three stores have different strengths, I go to them for different things, and they are organized differently.

Enactment of a script will vary according to many factors, based on the individual, the items sought, and the store itself. The choice of store varies as well, and may come about for practical reasons (it's on my way home) or for personal ones (I have a friend who always shops at Mr. D's, just because she likes it there).

Assignment 5: Grocery Store Organization

Date: 2/17/02

Bloomingfoods Co-op & Deli on East 3rd Street

I. Outline of organization by spatial arrangement and category.

NW corner (left of entrance):
Produce Section

North wall and opposite shelves, after Produce:
Purified Water machine, facing bottled water on shelves.
--then—
Sale items on outer wall, facing chips and salsas on opposite shelves.
Sale items here are usually juices, breakfast cereals, and/or chips.

NE corner:
Dairy case (including soy products and juice).

East wall, center:
Bulk section (grains, beans, powedered milk and various soy products, mixes, cereals and snacks). Islands of bulk spices, bulk oils, bulk nut butters.

SE corner:
1) Freezer section:
SE corner, plus island of frozen meats and end-case of frozen vegetables and fruits. SE corner continues around with additional cases of frozen breakfast foods and frozen meat substitutes.
2) Liquor section (mostly wine, against wall and on island. Some wine in Produce section.)

South wall:
Deli. Grab-n-go deli case; separate counter and cashier; hot-cold food bar. In center: cheese case and bakery displays; bagel bins. Coffee bar at foot of hot-cold food bar.
Faces north-south rows of vitamins, supplements, health aids and hygiene/beauty products.

Aisles: from North to South (left to right from entrance):
Note: each aisle features cases at each end for sale or special items. These cases face “out”, toward people entering the aisles.
--need to go back and take notes on aisles--

West wall:
Bathrooms. Exit Doors. Dining area.
Racks of cards, clothing, books and booklets (do they still have the clothing?).

Help desk.
Entryway. (Produce)

Cashiers
In SW quadrant leading to exit and dining area.

II. Analysis

The Bloomingfoods on east Third Street is organized to more closely resemble a conventional grocery store than the original Bloomingfoods off of Kirkwood. The original Bloomingfoods store is more "hippie" like, with wood floors, warm lighting, bulk items arrayed in large barrels on the floor, a single counter and cashier, and stairs to a loft area with additional items (no shopping carts here). In the 3rd St. store, however, the layout looks a lot more like a scale version of Kroger's. The building is rectangular, and all the merchandise is displayed on one floor. There are high ceilings with bright fluorescent lighting over light-colored vinyl flooring; ranks of cashiers at the front of the store; rows of aisles with signs to indicate contents; bulk foods are mostly displayed in shiny plastic bins; etc. The left-to-right organization of the aisles is similar to a conventional grocery store's as well.

This model for a grocery store was no doubt chosen deliberately in order to make the Co-op attractive to a wider spectrum of customers. The trend in audience for "natural" foods from penniless hippies to well-heeled baby boomers and professors represents the other side of that coin: not only would Bloomingfoods want to be attractive to people who weren't used to shopping out of wooden barrels, but their immediate customer base changed in its character and expectations.

The location and even the shape of the building also contributed to this design for Bloomingfoods. Adjacent to the Mall, across a parking lot from K-Mart, it made sense for this store to look more or less "like" other stores, but with specialized content. The physical shape of the building, a one-storey rectangle, also would tend to enforce a certain conformity with the "grocery store" prototype (although they could have chosen different lighting and flooring). Also, with this move to a larger facility Bloomingfoods could offer more items and more quantities of each item than the more traditional, smaller co-op or natural food store prototype provided. Physical factors such as size, shape and location are just as much a part of the design process as categorizations and intended audiences.

Assignment 6: Yahoo

Date: 3/6/02

The Yahoo! Directory is divided into 14 main categories, which are arrayed alphabetically in two columns on the Yahoo! home page. Those categories include: Arts & Humanities; Business & Economy; Computers & Internet; Education; Entertainment; Government; Health; News & Media; Recreation & Sports; Reference; Regional; Science; Social Science; and Society and Culture. However, Yahoo! is more than its Directory, including e-mail, instant messaging, maps, a calendar service, topical Yahoo! Guides, and regionally specialized Yahoo!'s, among others. The Yahoo! home page harbors a myriad of vectors for exploration and activity. Yahoo! provides multiple "ways in" to similar pursuits, allowing for any number of strategies a user might employ (or simply stumble into) in scanning the page. Having noted this, however, I believe that the intent of this assignment was to focus on the classification system embodied in the Yahoo! Directory.

The decision to present the categories alphabetically immediately presents the problem of separating some topics that should logically be closer together. There is also no reference (that I could find) to the entire array: you can only find sub-topics by clicking through or by searching the Directory. Because Yahoo! is trying to meet users at whatever angle they might be approaching their goals, there is a disregard for such things as Ranganathan's Canon of Exclusiveness. It better suits Yahoo!'s purpose to place different connotations of a subject into different categories, the better to generate "hits"—thus resulting in multiple iterations of ideas. Users are directed to these variants via the @ symbol: appearing at the end of a sub-category link, it indicates that this topic is referenced in multiple places in the Directory. In most cases, the @ link takes you away from your current Category browse chain and across to the related Category (i.e., the @ link can point to both coordinate and collateral classes, as well as to classes in other categories both higher or lower than the user's current position in the hierarchy).

This cross-referencing system, while generally productive for the user, is not fool-proof, however. Someone browsing within one "aspect" of a topic may not find out that there is another section on Yahoo! containing additional pertinent links. An example of this is Sexology, a sub-category of Social Science, vs. Sexuality, a sub-category of Society and Culture. Someone browsing in Sexology is given links to the Sexuality topic; but the reverse is not the case. A person exploring Sexuality in Society & Culture will not readily be directed to Sexology in Social Science. It is also very easy to get "lost" in the Directory by following these cross-reference chains. Although the "path" of the current browse is always displayed, it can be difficult to get back to one's original path if the "detour" strays too far.

Each sub-category is displayed with gateways to other areas of Yahoo! (such as Yahoo!Reference) at the top, followed by Category Matches; Sponsor Matches (paid links); and Website Matches. Again, without a "map" to the larger structure the process of browsing can be hit-or-miss; and as with any

classification, the relationships between subjects that are expressed in the structure may not correspond with the way the user is thinking of his/her project.

The Yahoo! Directory has a hierarchical classification system, with nested classes organized in data trees. Its strong cross-referencing features add a faceted or synthetic "feel" to its use, however.

A Yahoo! Directory search on "classification" yielded interesting results. There were seven Category Matches, the first being to the NAICS codification of business and industry types. "Library Classification on the Web", however, came second, and probably represents the target of this assignment. Further down among the seven were direct links to Indices to Web Documents>Dewey Decimal Classification (fourth), Indices to Web Documents>Library of Congress Classification (fifth), and Library Cataloging>LC Classification (sixth). 223 Website Matches appeared on the search results for "classification"; the official "Library of Congress Classification Outline" page is #3, "LC Classification Schedules and Manuals" is #7, and the "Dewey Decimal Classification Home Page" on the OCLC site is #9. The home page of the Classification Society of North America, which would relate more broadly to our subject matter for this course, is #10. Other hits in the top 10 websites include the Classification And Rating Adminstration (hit #1: apparently the advocacy organization behind MPAA film ratings); the Standard Industrial Classification system (a predecessor to NAICS); the Structural Classification of Proteins; and Noun Classification in Swahili.

"Library Classification on the Web" appears in the Directory along the path Reference> Libraries> Professional Resources> Library Classification on the Web. Interestingly enough, the only two sub-links for this topic are "@" links: Dewey Decimal Classification, and Library of Congress Classification, already mentioned in the Category Matches above. Where do these links take you? Why, to Computers & Internet> Internet> World Wide Web> Searching the Web> Indices to Web Documents (under which one finds DDC and LCC)—of course. Yahoo! has found a rather roundabout way to direct browsers in librarianship topics over to where Dewey and LC have been assigned: rather incongruously, from a librarian's point of view, under Searching the Web. Yahoo! apparently wants the people who are learning how to search the web to find out about Dewey and LC, and also apparently expects them to be greater in number than the librarians or library-oriented explorers who might be pursuing the same topic.

Clicking on "Dewey Decimal Classification", one finds eight web links. Strangely, one gets only one link from the "Library of Congress Classification" sub-topic under Searching the Web (a site called "CyberStacks", which organizes its guide to web resources using LCC). However, as indicated in the Category Matches above, Yahoo! keeps an additional sub-topic, "LC Classification", along the following path: Reference> Libraries> Professional Resources> Technical Services> Cataloging> LC Classification (Yahoo!'s abbreviation of "LC", not mine). The LC Classification sub-topic features nine

links. Dewey, poor soul, does not have a similar "mirror" topic under Cataloging—thus demonstrating another inconsistency/incompleteness in the system (Ranganathan's Canons boom in the distance).

Yahoo! provides two additional lines of attack, beyond following the category and topic links. Across the top of the listings are links to subdivisions of your search, the first three narrowing the results (Categories, Websites, Web Pages), and the last two broadening them (News, Research Documents). There were no hits for "classification" under News, but Research Documents got 113,574! Many of those related to "classified" government documents. Finally, a box at the bottom of the results page offers "Related searches"—in this case, animal classification, plant classification, insect classification, classification of living things, and a single classification essay (on Dewey, as it turns out).

The search on "classification" led me to relevant topics with little difficulty. However, the separation of library classification links in the organizational hierarchy, between web searching and librarianship references, made navigation of the results somewhat awkward. There appears to be no way to collocate all of the references to LCC, for instance; you have to go to both places in the hierarchy in order to see all the links Yahoo! includes in the Directory. Finding these "hits" by browsing would have been cumbersome. I for one would not have looked under "Internet" first. With enough persistence I probably would have stumbled across it eventually; but browsing on a topic like "classification" is difficult, particularly when you are after just one aspect of it—i.e., the library side, not the biological or industrial side. Yahoo! is designed to meet as many interests at once as possible—which does not always cater well to specialized expertise.

Knowing from experience that my "categories" probably won't match the system's, I would normally not attempt to find a topic like "classification" by browsing, but would search instead—either on "classification" itself or on the larger category of something like "library science". Yahoo! in fact recommends this "general" search strategy in the search tips section of the Help pages. "If you're not looking for a specific web site," they say, "**choose search terms based on the general subject** you're interested in" (emphasis theirs). Following this strategy, though, I would have found myself in Reference> Libraries> Library and Information Science—where "classification" is not to be found, not even as an @ link. 44 pages await you on "Metadata", but there is no reference under Library Science to Dewey or LC. Hmmmmm. "Library and Information Science" is an underdeveloped topic on the Yahoo! Directory, and the topics that a librarian or LIS student would expect to find there have largely already been assigned to other places in the Directory structure. With a project as huge as Yahoo!, these things are bound to occur as the Directory develops over time. Are Yahoo!'s editors subject specialists, or just web folk? Do they ever re-examine their subject assignments or attempt to revise the structure, as Dewey and LC do? Hmmmmm.

I found "Library and Information Science", somewhat surprisingly, through a hit for the Classification Society of North America, which is housed under >Library and Information Science >Organizations. This let me navigate

up to the LIS section, where I found possibly the one link (aside from the CSNA site itself) which might have been of real use if I were seeking general information about the idea or principles of classification: ODLIS, the Online Dictionary of Library and Information Science. The dictionary includes multiply hyper-linked entries on classification; classification schedule; classification system; Colon Classification; Dewey; and LC (as well as the professional organizations relating to classification).

Assignment 7: Faceted and Enumerative Classification Schemes

Date: 3/6/02

Shera's seven properties of traditional classification schemes include Linearity, Inclusivity, Significance of Terms, Significance of Arrangement, Meaningful Differences, Uniqueness of Arrangement, and Infinite Hospitality. Faceted and enumerative schemes enact these principles differently, according to their respective strengths and weaknesses. As has been observed in our class sessions as well as our readings (primarily Hunter, Vickery, and Ranganathan), the strengths of faceted systems lie in the flexibility granted by the formation of compound concepts, which also allows for greater consistency and the retention of relationships between subjects; enumerative systems, on the other hand, lend themselves to simpler notational structures and more straight-forward relations between the structure of the system and actual shelf order. In general terms, a faceted system has a superior schedule structure for reference to ideas (in the catalog), while an enumerative system has a superior citation structure for reference to items (on the shelf).

Linearity refers to the placement together of similar items in a logical sequence. Linearity in the ideal sense creates a gradation of subject matter, so that one idea flows into its neighbor. This usually refers to items on a shelf, as an aid to browsing (as Dewey was so delighted to find that his system accomplished so handily at Amherst). In a well-built faceted system you are more likely to find linearity well expressed in the schedule order than in shelving. The combinatory nature of faceting makes it easier to avoid arbitrary assignments when a decision must be made about category membership, and also makes it easier to show relationships between category features. But then you have to make difficult decisions for citation order that may end up separating concepts that in the schedule are clearly related. Enumerative systems cannot easily accommodate the conceptual flexibility that a faceted system readily supplies; but the linearity it manages to achieve within its structure will more readily remain in the citation and on the shelf. With regard to linearity, there are very rarely easy or straight-forward choices.

Inclusivity indicates the ability of the system to encompass all areas of knowledge. This is a daunting task, and like so much in the nature of classification remains a Platonic Ideal not often realized in the familiar four dimensions. Faceted systems prove stronger on this front, since the combinatorial structure allows more easily for the creation of new categories and the expression of new or revised ideas. The best-designed enumerative system, though, which attempts to anticipate future developments or the discovery of overlooked categories by leaving "room" in the notational structure, will still suffer. More frequent revision is required in an enumerative system, as old categories come to be inappropriate or obsolete, and new categories must be added.

Significance of Terms means that the subject headings or descriptors in the system must be specific and unique, meaningful to both the classifier and the user. Again this is a daunting standard. Any classification system is going

to run into problems on these battlefields of uniqueness and meaningfulness. Due to the multivalent nature of natural language and individual usage (mental models!), not to mention individual research goals and strategies that may not have been anticipated by the system's designers, the best-laid plans of classifications and classifiers gang aft agley. Faceted systems once again benefit from their conceptual flexibility, and also from the "bottom-up" nature of their formulation. According to Vickery, faceted systems begin with terminology: the strongest keywords associated with a certain topic or domain. A properly designed faceted system will rest on well-chosen descriptors, based on user need. A carefully conceived enumerative system will of course aim for the same goal, paying close attention to the best structuring of subject headings; but it is designed from the top down, beginning with categories and ending with items. Moving from the general to the specific, an enumerative system is more likely than a faceted one to miss terms that are important to the user.

Significance of Arrangement points to an array of subjects that effectively defines and orders a category, and is meaningful to both classifier and user. Faceted systems, with their strengths in showing conceptual relationships, are more apt to fulfill this principle. A well-designed enumerative system can still do well here, but it is less in its nature to do so consistently across the entire system.

Meaningful Differences refers to differences between classes. Yet again, a faceted system, with its talent for elucidating conceptual relationships, is more likely to do well here. Even the best-designed enumerative system is going to have points where it simply cannot show relationships across categories (without aggressive cross-referencing).

Uniqueness of Arrangement means that there should be one and only one place for each document. This is a very stiff order, indeed, especially in the modern-day burgeoning of interdisciplinary interests. This standard may become less and less important as the very idea of content storage and retrieval changes in the wake of innovations in automation and search capacities. The tools we use have an effect on our attitudes and on our strategies and requirements. As was demonstrated in the Yahoo! assignment, user-defeating arrangements in the structure are eliminated (or nearly so) by effective search capability that collocates across classification levels. Arbitrary decisions, and rules to guide them, are necessary to create uniqueness of arrangement *within the system*, even when the material itself suggests multiple placements.

Both enumerative and faceted systems have to create "uniqueness of arrangement" through the application of rules. These rules are necessary for the formation of either system, to guide the judgments of classifiers. With enumerative systems this process may be less jarring, since the "top-down" approach from general to specific requires greater arbitration from the beginning than faceting. For faceting, the application of rules can seem at odds with the conceptual flexibility which is the greatest strength of this approach. At some point, flexibility must give way to judgment informed by rules, which themselves represent arbitrary decisions (albeit as well-informed and thoughtful as possible). The compound subjects of faceting, while allowing

fuller expression of a concept, still require compromises when it comes to uniqueness of arrangement. Decisions have to be made as to which partner in the combination comes "first", and which emphasis of usage should take priority. In either system, rules must be established to shape the final judgment call, such as the "PMEST" ordering of the Colon Classification. Ultimately, the same questions confront the designer of a faceted system as confront the designer of an enumerative system—for instance, does "steel wire manufacturing" belong with steel or with wire or with manufacturing?

Infinite Hospitality represents the idea that a classification system should allow for perpetual expansion and accommodation. This is closely related to the ramifications of Inclusivity over time, but warrants its own principle. You shouldn't "run out of room" in the classification, whether you are attempting to subdivide or focus a topic, or adding new subjects. Here again faceting has the advantage with its compound subjects. A faceted system is more easily customized to a particular collection or particular user group. An enumerative system often requires revision in order to expand.

One issue that Shera's Seven Principles do not address directly is that of notation. Ideally, a classification system should produce citations that are coherent, consistent, and not too long. Faceted systems often violate the last of these points, since the expression of combined facets can produce a lengthy notation that is difficult to interpret, to place on the book, and to relate to a shelving system. An enumerative system with some synthetic elements, like Dewey Decimal, is more likely to satisfy on this level.

In conclusion, it must be said that even the best system, most aptly conforming to the referenced material and to the anticipated needs of the user, still must be learned. You have to know the organizational structure, the rules of assignment, and the notational conventions before you can use the system effectively. This is especially true for classifiers and for the maintenance of the system—in other words, for the librarians. On the user side of the equation, this problem is less pronounced, depending on the hospitality of the system. A system that promotes subject browsing (whether in shelf order or online) and allows for concept searches can direct novice users effectively to desired material. Hierarchical systems tend to separate potentially related subjects, while faceting is better at bringing them together—within the schedule, at least.

As noted earlier, developing techniques in information retrieval begin to break down some of these principles. Jenn has pointed out through her experience with the Digital Library project that a flexible, well-thought-out faceted system, combined with sophisticated natural-language IR capacities, may provide the best foundation for future information organization structures.

Assignment 8: Natural Language vs. Controlled Vocabulary

Date: 4/27/02

Indexing by extraction and indexing by assignment each have their strengths. Their effectiveness can be evaluated by contrasting the advantages and disadvantages of natural language versus controlled vocabulary. Indexing by extraction derives its index terms directly from the document itself, thus allowing for natural language searching. A searcher can retrieve the document by using words also used by the author, which have been selected for extraction as index terms. Indexing by assignment applies terms to the document from the controlled vocabulary of an established indexing language or thesaurus. These terms will not necessarily reflect language actually used in the document, but instead indicate concepts found in the document that correspond to authorized terms in the indexing language. A searcher can retrieve the document by using terms in the indexing language that have been assigned to the document.

The natural language approach of indexing by extraction has the advantages of relating directly to the diction of the document, and of allowing the searcher to use his or her own words rather than learning (or stumbling across) a controlled vocabulary. Additionally, any new or innovative terms in the document are immediately registered for indexing and searching, maintaining the currency of the system without additional effort. The system is easier to set up, since no indexing language must be devised: the document itself provides the index terms. Indexing by extraction can also be automated, contributing further to the advantage of low effort and low expense on the part of the institution making the document accessible. These advantages each have flip sides that become disadvantages, however: the searcher's "natural" words may not match those in the document, even if the intent is similar; the searcher has to work harder to refine searches to make up for the work that wasn't done by the presenting organization; and words used in the document may become outdated, or may shift in meaning, thus "losing" the document for retrieval if the searcher doesn't think to use the "old" terms.

The most significant disadvantages of natural language stem from the ambiguous nature of language and the idiosyncratic nature of its use. A word that means one thing in one context may mean something else in another, or may mean different things to different people or in different disciplines. Synonyms and homonyms also cause problems, where several different words may be used to express the same basic idea, or where words may be spelled the same but mean different things. Different word forms can create additional difficulties, such as "teaching" vs. "teach", or "contribution" vs. "contributes" (not to mention singular vs. plural nouns). Natural language brings with it colloquialisms, variations in spelling and misspellings as well. With indexing by extraction, the searcher's language must not only match that of the author, but must include terms from the document that happen to have been selected for indexing. While natural language indexing and searching are initially easier

and less expensive, they can add up to frustration and failure in searches due to the instability of language itself.

Controlled vocabulary offers the primary advantages of regularizing and normalizing language use, and establishing direct, one-to-one relationships between specific words and concepts. Indexing is more precise, and searching (once the system is learned) can be more efficient. The indexing language also provides a conceptual structure, revealing vertical hierarchies and horizontal relationships between ideas. The thesaurus approach provides for "lead-in terms" as well, allowing the system to acknowledge alternative approaches to the same concept while still directing the user to the authorized term.

A controlled vocabulary, however, takes time and effort (and therefore money) to design and establish. It also cannot (for the most part) be automated, and relies on human judgment to assign terms accurately and completely in anticipation of all likely searches. Inconsistencies among indexers are notorious, widening the margins for error. A searcher may be using a legitimate term from the controlled vocabulary, and it may even apply to the document, but the term has to be assigned to the document in order for it to be retrieved. Another serious disadvantage is the currency of terminology: new terms must be deliberately added to the system, and the process of discovering and integrating the new vocabulary entails additional time, effort and cost. Controlled vocabularies inevitably lag behind terminological trends.

Natural language and controlled vocabulary each have their advantages and disadvantages. Natural language offers simplicity and ease (at least initially) both for the indexing institution and the searcher, but can be stymied by the variability of language use. Controlled vocabulary provides conceptual structure and eliminates linguistic vagaries, but requires greater effort to set up and to use, and also has difficulty keeping up with terminological change. Ideally, the two approaches should be combined, allowing for natural-language keyword searching while also including controlled vocabulary for more structured searches.

Assignment 9: "As We May Think"

Date: 4/27/02

1) indicative abstract
Bush, V. (1996/1945). As we may think. *Interactions*, 3(2), 35-46. [Originally published in *Atlantic Monthly*, 176 (1), 101-108.] Available at: http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm

States that research and discovery continue to expand beyond the ability of the individual to access and comprehend the published record of human knowledge. Proposes that technology can develop ways to aid the individual to continuously extend, store and consult that record. Extrapolates from current technologies to suggest machines that might aid in performing research, as well as accessing the record of human knowledge and making connections within it. –ama 4/27/02

2) informative abstract
Bush, V. (1996/1945). As we may think. *Interactions*, 3(2), 35-46. [Originally published in *Atlantic Monthly*, 176 (1), 101-108.] Available at: http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm

Finds that the volume and pace of publishing outstrips our ability to make real use of the material. Concludes that machines can aid in the endeavors of research as well as act as a memory aid in extending, storing and consulting the record of human knowledge. Recommends the development of technologies such as photocells, thermionic tubes, dry photography, microphotography and microfilm, electronic transmission, voice records and transcribers, and calculating devices, to enhance intellectual work and free people from repetitive, lower-level tasks in manipulating data and studying the world. Proposes a device called a memex to act as a personal library, allowing not only for storage and retrieval of documents but for making connections (called associative trails) between ideas in different documents. –ama 4/27/02

3) indexing by extraction, from the text (avoiding pre-coordinate terms)

- Record
- Human knowledge
- Extend
- Store
- Consult
- Memory
- Mechanical aids
- Machine
- Mechanism
- Mechanization
- Device
- Transmit

- Review
- Photography
- Dry photography
- Microphotography
- Microfilm
- Photocell
- Thermionic tube
- Vocoder
- Research
- Repetitive processes
- Logical processes
- Symbolism
- Selection
- Association
- Indexing

4) indexing by assignment (ASIS)

- Human computer interaction
- Interfaces (UF man-machine interfaces, human-computer interfaces)
- Information storage and retrieval systems
- Organization of information
- Knowledge
- Memory (human)
- Expert systems
- Knowledge engineering
- Knowledge representation
- Information representations
- Information access
- Information processing
- Information overload
- Technology impact
- Logical skills
- Information needs
- Records (UF machine-readable records)
- Links (hypermedia)
- Hypertext
- Associative relationships
- Computer systems
- Interactive systems
- Cybernetics
- Computer applications (UF automation)
- Microfilm
- Audiovisual equipment
- Optical equipment

5) Library of Congress Subject Headings (at least three)

- Human-machine systems
- Human information processing
- Information storage and retrieval systems
- Information resources
- Electronic information resource searching

Subject Heading "most appropriate:"

- Information storage and retrieval systems

6) DDC and LCC: one class each to match "most appropriate" heading above:

- DDC: 004 Data processing Computer science
- LCC: T58.5-58.64 Information technology

Note: I changed my "most appropriate" Subject Heading from "Human-machine systems" to "Information storage and retrieval systems" after pursuing DDC and LCC and discovering that "Human-machine systems" may refer to industry and ergonomics rather than the human-machine systems Bush proposes. I had to settle. ☺ For added interest, I double-checked myself in IUCAT to see how books on information storage and retrieval systems have been classified, and found (of course) some inconsistencies, but the most common assignment was to QA76.9 or thereabouts: Science, Mathematics, Instruments and machines, Electronic computers, Computer science. I leave my choice of "Information technology" above anyway.

Assignment 10: Image collections (U. Cal. Berkeley)

Date: 4/28/02

The University of California-Berkeley Digital Libraries Project houses several photo collections, and also runs research projects into "computer vision". These collections can be described in terms of their organization, their "ofness" vs. "aboutness", and their access interfaces.

Blobworld (http://dlp.cs.berkeley.edu/photos/blobworld) represents one application for computer vision, where software recognizes groupings of pixels in the picture by color, texture and location, and defines areas called "blobs". Those blobs are then searchable: click on the blob, define a few parameters, and the search engine retrieves pictures with similar blobs. The CalPhotos image collection (http://elib.cs.berkeley.edu/photos/about.shtml) is more conventional in its interface, with retrieval based on text searches of information attached to the image in the cataloging process. The Blobworld search engine retrieves commercial images from a set of Corel CDs; the images cannot be downloaded. The CalPhotos collection, on the other hand, consists of images donated by photographers on an ongoing basis, with standards of use defined by the donor.

Let us begin with Blobworld. The Corel images on which this project operates can also be accessed more conventionally through the CalPhotos interface (to be discussed next). What is special about this project, however, is the opportunity to search by physical properties of image constituents, rather than solely by text labels. Such a search can render some interesting results.

From the Blobworld site, the collection is organized in two different ways. Underlyingly, the materials the site works with are organized according to the Corel CD's that are the source for the images. 350 CD's with 100 images each make up the collection. Each image file has a number and title assigned by Corel, along with keywords describing content, also assigned by Corel. The CD's are not organized by content or theme, but simply by file number.

While CD content lists are available from the Blobworld site as a reference, they are not searchable as such and are not meant to guide the user. The images have also been divided by the Blobworld project into six categories according to content: Animals, People, Flowers, Ocean Scenes, Outdoor Scenes, and Manmade Objects. The user can choose examples from one of these categories as a starting place for a Blobworld search.

The Blobworld approach is one of "of", not "about"; pictures are described as being "of" certain things, and their meaning or implication is not addressed. These are pictures of animals, people, flowers, etc.—not pictures about family or about the ecosystem or about cultural documentation or about the scientific gaze (although the user is of course free to apply such interpretations and to seek images using such "aboutness" criteria).

All of the above is information one can glean through exploration of the site. The initial interface greeting the user is rather stark and uninformative about the content of the collection or how it is organized. The focus here is on the experimental blob-based search capacity; usability for finding desired

images appears to be neglected for now. This may be due in part because the commercial Corel images on which the project is based are not downloadable except by linking to Corel and paying a fee. A search cannot be made for "practical" purposes, the designers may have reasoned; for now, the interface exists primarily for testing of the system.

A navigation column on the left offers text links such as "List of images," "sample queries," "starting images," and "source code". The main body of the page provides a brief description of the Blobworld project, then offers links to "Try a Blobworld query," take a tour of a sample query, or read a technical paper about Blobworld. There is no immediate information about what kinds of images you'll be searching, although a sample photo of a wolf is given along with its "blob" representation. Although this is not explained on the page, if you click on the wolf picture or its blob image you will initiate a query with that image.

If you want to know about the images before proceeding, you can follow the left-column link to see not a listing of each of the 35,000 images, but of the 350 Corel CDs they come from. The CDs are listed in order by Corel's item number for each disc, along with the title for the disc. There is no grouping by "subject", no sorting function, and no search capacity on the page. By scanning the list you get the idea that there are pictures of sunsets and birds, etc., but this is not a very useful or effective presentation from the point of view of a searcher. You also cannot access the CDs from the page. An odd feature is offered, whereby you can have the CD number display alone—with no titles. Why is this option available—what purpose does it serve? I found this page frustrating and mildly pointless. Perhaps it was required for disclosure purposes of the research project: these are the materials we used from Corel (sigh). As stated earlier, this project is research-based, and not yet entirely user-friendly.

From the main page, it is not immediately clear how you ought to begin. Do you just jump in with a query? Should you take the tour first? Nothing says, "Start here". One of your options is to "try" a Blobworld query; another is to take a "tour" of a sample query. If you decide to take that tour first to get oriented to the system, you may be thrown off base before you start. The tour starts not with the first page you see when you begin a query, but with the results of that initial screen, after you choose a category. The tour itself is interesting and reasonably well designed; but if you proceed from the tour to "trying" a query, the first screen of the query process will be a new one on you.

The first query screen shows example images, and instructs you to click on one to begin—or click on one of the category links to find another image (the categories being those I listed previously). No blobs are displayed on this page. Access is therefore provided first by the image itself. When you choose an image, the next page shows the image, plus its "Blobworld" version. There is also a Keyword text box, and a Submit button. The bottom of the screen gives you instructions for how to proceed.

Your next point of access is either one of the "blob" areas within the blob-ified image (chosen by clicking on it), or terms typed in the Keyword box,

or a combination of the two. When you choose a blob, you can then weight different qualities of the blob for determining the parameters of your search: the importance to you of the selected region; the importance of its various features, weighted individually (color, texture, location, and shape/size); and the importance of the background. Radio buttons let you select "Not", "Somewhat" or "Very" for each option. Access to the images is provided through the images themselves; through their physical "contents" as determined by Blobworld's software and as weighted by the user for importance; and through optional keywords.

Addition of the keyword search makes things a little interesting. It is possible to enter Blobworld and perform purely a keyword search, without taking advantage of blobs. However, the designers discourage you from doing so—in fact, discourage you from using text at all in your search—with a warning against it. This warning link brings you a double-edged caveat that keywords 1) search only the tags assigned by Corel for title and keywords, so your search will have to match their assignments in order to be fruitful; 2) are fraught with peril because Corel's tags are fraught with typographical errors, so that even if you're ostensibly using the same word as they are, you may still miss because they spelled it wrong (no mention is made of *you* spelling it wrong).

I found this warning a little interesting, primarily because combining keywords with your blob-based search is the only way to focus your results on the "of" you're after (assuming you're after an "of"). If you want to be sure that your blob search on a picture of a wolf returns only (or mostly) other pictures of wolves, you have to use "wolf" as a keyword (and even then you might get surprises). Otherwise, the wolf-shaped blob is just as likely to "match" with a section of a dirigible, or a bit of sky, or moss on a rock, or the corner of a building, even with carefully chosen weights of image characteristics. While this can be fun for exploration of the system's capacities (which is the current raison d'etre for the site), it could be frustrating for someone with a more specific search in mind. Of course, if all you're searching for is pictures of wolves, a more conventional text-based search interface (or a keyword-only search) may get you what you want anyway (even with typos in the indexing tags).

Interestingly, given this warning, using keywords is the only way to access the individual image record from the CD. A blob-only search gives you images and blobs; the CD image number is shown, but that is as far as it goes. Incorporate keywords in your search, and the image number becomes a link to the CD info, taking you to a new page displaying the photo and its blob image along with the identifying tags provided by Corel. Click on the blob image on this image info page, and you can begin another search. This is a useful and informative page, and one wonders why it could not be made available to the faithful blob-only users who follow instructions and heed warnings.

Another surprise in relation to the keyword warning is that if once you use a keyword in your search, that same keyword remains in force for subsequent searches—even though it does not display in the text box when you

search again. You only know it's there because of the way the results are displayed (the keyword is mentioned in your summary of hits, and the image number is a link). To remove the keyword from subsequent searches (particularly if you are no longer pursuing that particular "of"!), you have to click the "Clear" button (which seems silly when from the user's perspective the text box is already empty), which resets your search.

If the keyword warning was provided primarily to prevent people from coming to Blobworld and doing only a text search, it would have been easy enough simply to program that option away, disabling the Keyword box if a blob is not selected (and making this clear in the instructions). Instead, a somewhat misleadingly alarmist warning is given. Hmm.

Following the paths down which a blob search can take you, it is easy to lose any sense of content-based organization. Instead, you watch for what the blob you chose will "turn into" in the next set of results. Blobworld is not about the organization of the collection, but about testing the experimental blob-recognition process. If you are an artist or designer looking for specific colors, shapes or textures rather than more conventional objects or models, this kind of search could be highly productive. It's a fascinating experiment to play with.

Navigation is made somewhat complicated by the decision to open each search in a new window. Each time you click the "Submit" button for a search or a refinement of a search, you are in a new process, which is a good way to get the user rather lost if they keep playing with the system or keep trying to get something specific. No navigation is provided for returning to the main page and starting over, or for starting a completely fresh query, either. Instead, the queries are self-propagating, even while being disconnected. In some ways this is useful—you can click between windows to remind yourself of the steps in your journey among the blobs—but it also takes up system resources. Depending on the strength of your computer, too many windows and the system starts returning blank pages rather than search results (either that or adding the keyword "bansai" caused some kind of error).

Without any navigation "home" or to collapse the current chain and start fresh, the user must return to the "root" window and use the browser's back button in order to make a new start. The left-hand column featuring details about the project is also absent once you begin a query. That column (including a "Blobworld home" link) is present on the first page of a new query, where you choose your example or your category, but disappears once you being querying. If you want to consult any of that information you have to stop what you're doing and use the "back" button to return to it. It would aid the user greatly to include navigational features on the query and results pages.

I understand why usability is often applied as an afterthought; for scientists and programmers, the first priority is to get the silly thing working. However, this makes either for a major overhaul of the presentation later (essentially re-writing it), or leaves the final product very rough around the edges from the perspective of the user (or, usually, a bit of both). Very little effort would have to be expended to make the Blobworld entrance and navigation more assured and more clear to the user. This is a big part of what I

do in my job as "website testing specialist" (the least favorite part of my job), and I'll freely admit that it can be frustrating as all get-out to have to point some of these things out *after* an interface has been designed.

The CalPhotos collection is accessed through a more conventional text-based system. From the main page you can choose to query CalPhotos as a whole, or focus on one of five CalPhotos subdivisions: Plants, Fungi, Animals, Landscapes, and People and Culture. Four other collections beyond CalPhotos are also offered on the same gateway page: California Dept. of Water Resources, Africa Photos from the California Academy of Sciences, the Corel Stock Photos on which Blobworld is based (here accessed via a text-based query form), and Aerial Photos of the Sacramento River Delta region, also from the California Dept. of Water Resources. A menu in the left-hand column offers links to Frequently Asked Questions; References; Usage and Permissions; Contributing Photos; Contributing Annotations; Photographers; Organizations; About the System; Source Code; Database Schema; Computer Vision Research; and History of the Image Collection. Each category has a brief description next to it, including a link to view a sample from the collection. Once again, these collections are oriented around "of" rather than "about": pictures are "of" a species of Madagascar chameleon, for example. As for so many things in the world, the "about" is up to you.

Access to these collections is provided through a query form. This form lets the user make text-based searches on fields such as name, location, type of picture, photographer, and even color. Combo boxes provide drop-down lists (with defaults of "any" or "none") for all the fields except for name and location, which are free-text entry. There is no option for a "keywords anywhere" type of search; some structure to your query is required, although it is made easy for you with the layout of the fields. The form then submits a SQL query to the Informia database housing the indexing material on the images, and the results are presented as a web page. If you leave all fields blank or unselected, your query will let you browse the entire collection you are searching. You can select an option to display the Review status of the photograph (whether or not the information provided by the photographer has been verified as to species name). Another option allows you to display only text information about items without the image itself.

With the exception of the Corel Stock photos, the images can be downloaded, or linked to using a custom query encoded in a URL (instructions for doing this are provided on the website). Each picture's data includes at minimum a taxon name for the item photographed (in the case of plant or animal), the photographer's name, and a unique identification number. Photographers may also provide location information, date of photograph, their common name for the item photographed, etc. Search results can also include links to zoological or botanical database entries online (some of which send you to defunct or moved pages), and you can also choose to view all photos of this animal, plant, or habitat. There is also a link to add your own comments to an image (I didn't try this, so I don't know how such input is moderated; password control is likely).

Returning to the exact page we were given as an entry point for this assignment can be very difficult. Image records provide navigation links back to their home collections, but not to any kind of hub such as the URL for this assignment. Once you get to the home page for a particular collection, navigation and presentation may vary, and your search options may change as well. Image records do not provide navigation back to the search query form, either; as for Blobworld, you must use your Back button or return to the collection home page in order to begin a new query. There appear to be multiple points of access to these collections, which can be a good thing—so long as access is also coordinated and navigation is consistent. The webmasters for the Berkeley Digital Library obviously have an enormous job, assuming there is any central coordinating of these various projects. Users of the collections would certainly benefit from systematic, consistent navigation design for the various collections, with more clearly labeled central hubs for access to the system.

Obviously, organizing image collections and making them accessible to searchers is a dauntingly complex task. The Blobworld project points in some interesting directions for the future of image searching. The CalPhotos collections offer a versatile search interface and a wealth of material, but some confusing navigation; and as long as identifying information must be left to the donor, inconsistencies are inevitable (of course, they're inevitable with librarians, too, but will probably be more systematic and predictable if done by professionals). I believe I may have taken this assignment a bit too seriously (?), but with my temperament and my "day job", I can't resist thorough exploration of a site.