I would like to present here for consideration and comment two initial explorations into the use of verbal descriptors for aboutness of image documents. I would also like to comment on the process of elicitation of those descriptors; then propose a larger study and solicit input on the conduct and refinement of that study. A major portion of this text appears in Explorations in Indexing and Abstracting: Pointing, Virtue, and Power, being published by Libraries Unlimited, and is used with their permission.

The difficulties imposed by the literary model on representing image documents caused my research efforts for a long time to look at structural, native element descriptive methods. Recently, a conversation with M. E. Maron of fifteen years ago came to mind and urged investigation of a word-based method. That conversation turned on the difference between subject and aboutness. This distinction is similar to Wilson’s distinction between topic and function.

While discussing the inappropriateness of words as descriptors of images because of the very different means by which the two sign systems operate, Maron suggested that surely, no matter
how a viewer interpreted a video document, s/he could say something about it. That something would likely be a reflection of the viewer’s reaction. Thus, no matter what relationship that reaction bore to the author’s intent or an indexer’s conceptual tag, it could be said that that something represented the document’s aboutness or that user.

The images in Figure 1, together with their accompanying descriptions, point to a major problem in the representation of any sort of documents. Different users may well have very different notions of what the document is about (for example: Maron, 1977; Robertson, 1979; Wilson, 1968) This highlights the access problem for users who must depend on the judgment and coding of someone else.

Photographs Are Not Words

Photographic images are not words. Photographic images are usually very specific representations made at particular moments, of particular objects. Words are general representations. Pictures are made more general by adding more pictures in a sequence or collage. Words are made more specific by grouping them with other words. Figure 2 illustrates this point.

Figure 2
We can say that photographs help to make document representation issues more obvious because of the very different ways in which pictures and words work. Word texts can be described with elements directly from the document and similar to daily speech acts. Thus the possibility for confusion of elements with topic and topic for aboutness runs high.

Likewise, word texts have clearly segmented elements set within rule bound structures. We can, for example, say that a particular word is a noun and because of its place and the places of other words, that noun is the subject of a sentence. At one level, then, we can determine topical characteristics of a text with some ease and surety. There are, of course, many caveats to such an approach. The meaning of a text for any particular user is heavily dependent on that user.
Image texts are not constructed in a manner which allows easy demarcation of elements or rules for extraction of a subject. Photographs are, in a sense, made by sampling a very broad band stream of data. They are analog representations with very fine gradations from light to dark. They present no easily discerned noun/verb analogs. There is no rule for translation of a whole image or any of its parts into words. (Novitz, 1977) The old phrase “one picture is worth a thousand words” speaks well to the high bandwidth of communication that is possible with image texts. However, there is no saying just how many words or just which words are required to describe any individual picture. The word document has easily discernable units and clusters of units of meaning. Photographs do not.

Humans have brains which are uniquely suited to visual information. Nearly fifty percent of the neocortex, the “higher”, primate portion of the brain, is devoted to visual processing. (Fischler & Firschein, 1987) We seem to be very good at pattern recognition. So, when we say that pictures cannot easily be translated into words, there is no implication of inferiority of images as a representation medium. As multimedia systems burgeon in many fields, the issues of image representation become more vexing and more compelling.

Pictures represent the object/event space in a manner fundamentally different from words. (O’Connor, 1985) In turn, representing pictures with words is a vexing challenge. Yet, people do, in fact, represent pictures with words. If you ask someone what a picture is about, they can usually say something. Reactions to the lantern slides pictured both above and below indicated that the “something” is often not just (or “even”) the object or set of objects in the image. Variety of potential usage generated a variety of conceptual descriptions. Choice among synonymous terms or level of specificity are not the only issues.

Representation of images by words becomes even more problematic when we consider the issue of generalization. The words “elephant”, “sheep”, “horse” can be generalized to “animals”. We have verbal representations of taxonomic relations. What though would we do with a photograph of an elephant, a photograph of sheep, and a photograph of a horse, as we have in Figure 3? Is it adequate to simply combine all the photographs into a collage? Are combined pictures really a better solution for some circumstances, since the word “animals” could really include many things besides the elephant, sheep, and horse?

---

**Figure 3**
college
Subject indeterminacy causes search failures within the realm of word based documents, in which both the documents and the representations are words. How, then, are we to represent photographs with words and expect successful searches? Abutness presents a challenge to which we have already alluded. If we have a difficult time avoiding indeterminacy in word representations of word documents, how can we possibly expand the number of conceptual tags with image based documents? While images have been in use for millenia, it has only been recently that any large percentage of the population has had the ability to make and use images. This complicates our question because there is no strong background of accessible visual literacy on which to construct picture based representations.

**Initial Explorations**

Two case studies of uses of photographs provide a different avenue of approach to the representation of documents. The first case involves a chance discovery of some antique lantern slide images, while the second is based on PhotoCD technology. The two cases span the use of photographs in educational environments, from the late nineteenth century to the present; and they both point toward an enriched mode of representation.

During the renovation of the administration building at a small university on the Great Plains, several small wooden boxes were discovered in the clutter. A few were salvaged because of their attractive appearance. One faculty member noticed that each box contained glass lantern slides and attempted to obtain as many boxes as possible. Approximately fifteen boxes were eventually located.

Each box contained approximately one hundred slides. Each is a sandwich of:

- two sheets of glass--.0625 in. x 4 in. x 3.25 in
- a piece of roll film, typically but not always, 2.25 in. x 3 in.
- masking material of various sorts
- tape bindings

The physical condition of the slides varies from excellent to poor. Many show no signs of wear or damage, while others have cracks in the glass or problems with mold growing. The subject matter of the necessarily haphazard sample of slides ranges widely. A partial list of the topic areas includes:

- hand tinted copies of engravings of the Aeneid
- portraits of writers and artists--Renaissance to late 19th. century
After these boxes of antique slides had been rescued from the brink of demolition, they sat unattended for several months, serving mostly as conversation pieces and paperweights. On occasion a faculty member would come across some of the slides and think of a way of using some of the images in teaching or research. Some of the images were unavailable from more standard sources. Since there are no projection facilities available on the campus for such slides, use was limited and interest did not turn to action.

By chance a few of the lantern slides were brought to the room where a digital analysis of video images project was underway. The addition of a home video camera to the computer imaging system enabled input of digitized images of approximately twenty of the lantern slide images.

On a casual, ad hoc basis various faculty members and graduate students called up the images on the computer and were uniformly pleased with the results. Several uses for the images in different courses and departments were conceived. Some of these included:

- source for stage settings and costuming
- lecture illustrations in history, classics, art history, English
- source for image fragments in video on collapse of Rome
- background images in desktop publishing
- comparison of artifacts with paintings of classical scenes
- hypermedia stacks for study of Shakespeare and antiquity

As people from various disciplines made comments and suggestions they also began to realize the need for some access system. Several also pointed out that a list of descriptors suitable for people in widely differing fields would have to be long and multifaceted. The stage dresser seeking an image of Hawthorne’s home would be seeking different aspects than would an English literature student, or a professor of architecture. The vocabulary of these differing users would also be quite different.

**Aboutness**

The very different reactions of these casual users brought to mind the conversation with Maron regarding verbal description of user reactions. *Aboutness* is the term we will use to distinguish functional representation from mere description or application of a topic.
We can say that aboutness is extra-descriptive. It is likely to be generated, at least in part, by the subject of a work, though it may be that a secondary element to one user will be a primary element to another. Yet it goes beyond that to include “What this means to me.” Aboutness is the behavioral reaction of a person to a document. Each patron may have a different experience with the same document. All of the elements we have discussed earlier on will come to play in the personal reaction to the subject elements. We might say that aboutness has an adjectival component in addition to the noun.

We can imagine the patron looking for “something cheery for springtime”, or “something depicting passionate commitment”, or “some images showing harmony”, or “something that makes me feel good.” We may say, then, that aboutness is, indeed, descriptive. It describes the relationship that holds between a user’s knowledge state and the physically present document.

Movie critics provide a good example of aboutness judgments. When some critics rave and others pan, it is not because they have seen different physical texts; rather, all the technical knowledge, topical knowledge, emotions, and beliefs of each critic are being engaged in the construction of a reaction to the physical text. Viewers may come to realize that their own complement of knowledge and belief and emotion structures more closely resemble one reviewer than others, so that the reviews of that critic will become surrogate aboutness judgments for the user.

Community Memory Interface

A word based system for describing the aboutness of pictures can be constructed by changing our model of where the act of representation takes place. Typically the rules for representation are established by some external agency - OCLC, technical services, Library of Congress, etc.. What if we were to re-establish the point of representation activity as the patron group?

The digital environment enables keeping track of large amounts of data. The storage and manipulation capabilities of a computer could substitute gathering and ranking user-generated descriptors for the typical mere storage of agency generated descriptors. Such an approach offered potential for:

- accumulation of as many descriptors as users thought appropriate
- accommodation of multiple functional concepts
- accommodation of multiple levels of specificity
- multiple terms for same object or concept
- user determined descriptive terms
- multiple formats of descriptive terms

There are, of course, also challenges:
A community memory interface to a collection makes several assumptions. It assumes a new relationship between the interface and the users of the system. The users will be contributing to the system, in a sense customizing it, nurturing it, teaching it. To illustrate this idea, imagine a recent graduate with a degree in library science beginning work at a reference desk. The new reference librarian, the interface to the collection, knows the documents, but is a blank slate regarding patrons. However, after a time, as clients come into the collection and ask questions, make their likes and dislikes known, and discuss their areas of need, the librarian will develop profiles. These will include the idiosyncracies of the more frequent clientele. The user profiles will enrich the librarian’s ability to select documents not only by topic, but also by all those attributes that contribute to “what it means to mean--how it suits my purposes.”

The interface is nurtured and enlarged and elaborated. Few patrons would expect the new reference librarian to be as facile and knowledgeable of individual representation schemes, as a librarian on the job for a year or more. So too, we may imagine a digital system which gathers input from users and grows and becomes more elaborate in its representational capabilities.

Such an interface also assumes that at least some of the patrons will, upon occasion, be willing to take the time and effort to contribute to the system. The imagined community memory interface for the picture collection assumes only the most minimal representation of pictures at first. Patrons may have to do random searches. A picture is found that is desirable for whatever reason, a request box will appear on the computer screen asking if the patron would care to add subject headings or comments about the picture to the access system.

As more and more use is made of the system, many images will begin to accumulate descriptors. Some images, though, may accumulate few or none. This will reflect the needs of the community using the documents. The patron who might be served by an image with few or no community tags will still have the option to browse through the images not yet labeled. Figure 4 diagrams the basic interface.

---

**Figure 4**
tension * anticipation * anxiety * anger
ruggedness * determination * courage * disgust
called * tough & rugged * unfeeling * resolute
determined * another era * Kansas * strong
facing a challenge * needs a bath * disappointment
dry & dusty * solitary * isolated * peaceful
lonely * peace * ownership * pride
independence * desolate * lovely * relaxing
barren & sad * boring * Arizona * vast
Indian reservation * beautiful * Colorado
The sample size is still too small to make significant generalizations about variety of concepts and variety of depth or level of specificity. Table 3 presents the images and data for four of the pictures and demonstrates considerable variance. Indeed, some images elicited descriptors which are nearly opposites. For comparison purposes, Table 4 presents subject headings in the manner of Library of Congress Subject Headings, applied to each of the test pictures by an art and architecture slide librarian.

Table 4
Mountains - Idaho * Ranch Houses * Trailers

Building Stones * Building, Brick * Blinds
Subsequent Exploration

The above suggests that pictures are not words, but words can be used as representation tools, especially if the construction of those tools is put into the hands of the users. The rules for highlighting and the methods of coding are made manifest to the users because they made them. Of course, this assumes either a certain homogeneity of users or means for a patron to select search terms applied by users with some particular profile. Now, it may turn out that if there are enough users of the system, no matter how heterogenous they may be, several small clusters of very different types of description may develop for some pictures. Many patrons, then, would have available representations constructed by members of the microcommunity to which they belong.

Aboutness, in the functional sense we have used, is a powerful representation because it directly includes the user’s knowledge state in the representation process. Evidence so far suggests that a community memory interface is one method of integrating aboutness into the retrieval process. It may just take some time to engineer the dynamic, learning system required for the implementation requests based on aboutness.

More substantial exploration revolves around putting a larger number of images before a larger and more heterogenous user group. The current plan is to construct a CD-ROM with one thousand images. The images are a random selection from the files of a team of media producers. It includes travel images, family portraits, animals, sports, flowers, birds, and many other subjects; some are photographed in what might be termed “standard” modes, while others are rendered “artistic” or “idiosyncratic” modes.

This will be placed in several different environments - small public library, academic library, middle school classroom, classrooms in different academic departments, etc. We are currently soliciting suggestions for appropriate sites. We will be looking to find degrees of overlap between user groups, or patterns of lack of overlap, together with further substantiation of the preliminary findings of broad ranges in terminology, levels of specificity, and aboutness. Again, any suggestions on the most efficacious manner of pursuit, as well as any other comments on the enterprise would be most appreciated.

References


• Martin Fischler and Oscar Firschein, Intelligence: The Eye, the Brain & the Computer (Reading, MA: Addison-Wesley, 1987).