The Visual Thesaurus in a Hypermedia Environment:
A Preliminary Exploration of Conceptual Issues and Applications

Matthew Hogan
Center for Science and Technology 4-330
Syracuse University Syracuse, NY 13224

Corinne Jorgensen
Center for Science and Technology 4-290
Syracuse University Syracuse, NY 13224

Peter Jorgensen
Interactive Publishing
Hamilton, NY 13346

Visual thesauri and their applications to collection management incorporate a complex mix of political, economic and systems design issues. This paper describes and compares printed visual dictionaries and current state-of-the art visual retrieval systems. We critically examine the assumptions upon which such systems are constructed, and explore some ways in which images can be used in a text-free information system, and propose some areas here further inquiry is needed. We also present a prototype for a hypermedia visual thesaurus. Those who catalog objects are just now beginning to recognize the inadequacy of verbal language as a means of recording descriptive information on material culture or the physical world in general. The use of non-verbal representations has great potential to address the limitations of text-based taxonomies. The use of authority control and thesauri are an essentail part on any successful information sysem. In this paper, we have briefly described some of the implications of a similar process for visual media.

Introduction

In a recent article in *Leonardo*, Lawrence E. Murr and James B. Williams addressed the issue of visual versus textual information in education. Educators and librarians for centuries have relied on the printed and spoken word as the dominant method for transmitting knowledge. Information specialists have continued this trend by application of text-based information retrieval aids, such as index and thesaurus construction and keyword searching.

Murr and Williams discuss recent research into the brain's functioning which has led to the model of the left-brain right-brain dichotomy.[1] In this model different functions of cognitive processing are located primarily in the left or right hemispheres of the brain. The brain's left hemisphere seems to be linked to language processing, and is well-exercised by the overall emphasis on speech and text in education and in information systems. The brain's right hemisphere, which handles spatial reasoning, symbolic processing, and pictorial interpretation, has been ignored or at best has received much less attention in system design. Only recently have researchers begun exploring such concepts as "navigation" (a right-brain function) in information systems.

The development of cinema and later television and the present-day expansion of consumer electronics with the VCR and camcorder have made it obvious that people have a strong affinity for images. With the development of graphical interfaces and the rapid integration of computer technology with image storage media such as the video laser disc, textual representation no longer has to be the dominant paradigm in education or information retrieval.

The current emphasis on text needs to be balanced by increasing attention to graphic formats and visual interfaces. Many current information systems used to access images simply transfer the text-based methods of information storage and retrieval to computerized systems. Because the multiple aspects of visual access to images has long been neglected its development is still woefully inadequate. As statistical computer methods for image retrieval become more powerful a new emphasis on research in the multiple aspects of visual retrieval are needed. The interrelationships between text and graphics need to be thoroughly explored, as does the creative possibilities in visual-based retrieval systems. At present more and more powerful computerized retrieval systems are being developed that, according to the left-brain/right-brain model, still use only half of the processing power of the human brain. Additionally, learning research shows there are many different ways in which human beings learn; text-based methods fail to develop all but a few of these.

Visual thesauri and their applications to the museum community incorporate a complex mix of political, economic and systems design issues. In this paper we describe and compare printed visual dictionaries and current state-of-the art visual retrieval systems, critically examine the assumptions upon which such systems are constructed, explore some of the ways images can be used in text-free information systems, and propose areas for further inquiry. We looked at systems which utilized images as an aid in information retrieval, and discuss three that we feel are representative.

We are presently engaged in the development of a prototype visual retrieval system which will allow further exploration of conceptual issues presented here. The system, operating in a limited domain for which a visual dictionary already exists, allows manipulation of the visual object as a search mechanism, as well as the traditional textual access points. The domain chosen for this project is leaf identification.

We begin with some background material on images and image processing in general.

Human Image Processing

Reed discusses a variety of evidence suggesting that visual images and visual codes play an important role in human ability to perform many cognitive tasks.[2] Images preserve the spatial relations among objects in a scene or the features of a pattern. Experiments have shown that people can use either a visual code or a verbal code to retrieve information; the choice is dependent on the nature of the task. When a task emphasizes spatial information, people can scan an image in a manner which preserves the spatial information. In contrast, if the task requires the retrieval of verbal information, scanning a list associated with spatial information but ignoring it, is more efficient.

More importantly for information retrieval, visual images make it possible to simultaneously compare all the features of two patterns. Therefore information is matched in parallel. This is in contrast to features described verbally which are not all accessible at the same time and must be compared serially. Matching has been demonstrated to be relatively fast and independent of the number of relevant features only when the initial item is a pattern, and not a verbal description.

Visual images also make possible mental rotation of patterns until both have the same orientation. It has been demonstrated that time to match two images differing in orientation increases linearly with the number of degrees by which they differ. A computer can rotate images, often much faster than the brain can, thus they have the potential to facilitate the process of visual pattern matching.

Visual images play an important part in memory and learning. Research has shown that people remember pictures better than concrete words, and concrete words better than abstract words. Similarly, the formation of a visual image is much easier for concrete material than for abstract material. Words vary in how easily they can be translated into images however, and the successful use of imagery depends somewhat on the nature of the material that has to be learned. The imagery potential of words is a more reliable predictor of learning than the association potential of words. High-imagery words are easier to learn than low- imagery words, but high association words are not necessarily easier to learn than low-association words. Even when subjects are told not to use imagery, high-imagery words are easier to recall than low-imagery words. This suggests that people spontaneously generate images whenever they can. Dual-coding theory explains the usefulness of visual imagery in recall by suggesting that a visual image provides an extra memory code which is independent from the verbal code.

Current research has supported several of the long-standing rules-of-thumb concerning use of images as memory aids. Images need to be of a certain size to be effective, and the use of distinctive images and loci are important. There are several limitations to the use of visual images as memory codes. While a person may be able to visualize an image, the retention of details is more problematic, and suggests that visual images may be incomplete, and at least partially based on expectations or schema. Other results from tests requiring the identification of parts of a pattern indicate that most retained images are highly structured,

somewhat vague, and difficult to reorganize. While visual imagery may not be appropriate for every task, the knowledge that people use visual as well as verbal codes to organize information deserves wider recognition and further research into how these processes could be capitalized upon in information retrieval generally, and in image retrieval specifically.

Museums and Images

Artifacts, objects, and material culture are the purview of museums, which classify and store their collections, as well as to make them available to scholars and the general public for education, enrichment, or research. At present, collection management tools used by the museum community rely on textual representations, while treating visual information or representations of objects as secondary. Collection management systems are not designed to deal with images as primary sources and so reinforce a text-oriented way of thinking about information retrieval. This includes but is not limited to, the interpretation of what the object is about, and what information carries the "aboutness" present in the object or its visual representation. Disciplines such as art and archaeology, which are actively involved with developing indexing and classification schemes for physical objects, are a source for detailed discussions on the problems and possible ways of indexing visual media. Even in non-object oriented fields the difficulty of developing non-ambiguous, consistently applied subjects and descriptive terms is widely recognized.

There is no universally accepted taxonomy, subject authority, or vocabulary for describing material culture. While over the last ten years several projects, such as the Art & Architecture Thesaurus (AAT) and Chenhall's Nomenclature, have developed hierarchical lists of descriptive terms, individual curators, archivists and librarians continue to develop their own indexing vocabulary and syntax. Standardization among systems which do exist is increasing, but, the characteristics of "aboutness" across and within institutional catalogs remain inconsistent and often difficult to determine. There are a number of problems associated with this approach, not the least of which is the difficulty a naive user has in easily accessing images, and that subject terms may influence subsequent interpretation of the image.

Karen Markey has described current access methods for collections of visual images.[3] She found that visual collections are accessible primarily by secondary subject matter, the overall theme or concept expressed by the image (such as *Madonna and Child*), rather than primary subject matter (the actual objects comprising the overall image). A major problem with this type of access is the unpredictable loss of a great deal of information about the image. Markey suggests the use of primary subject matter, a listing of preiconographical descriptors in an alphabetical, keyword in context format. This is the approach of the *Library of Congress Thesaurus for Graphic Materials: Topical Terms for Subject Access*, (Betz). Yet even this way preserves the inconsistencies of word meaning and problems of synonymy found in any text-based indexing system, in large part due to human error.

In a review of the literature on index creation for visual databases, one of the points mentioned time and again was a desire for a flexible search mechanism which would allow the

addition of special visually oriented search capabilities,[4] a feature which is currently not possible with the MARC record format for pictures. This **implies** that there are considerable limitations for those attempting to create an indexing system for images which will address the needs of users in a MARC based system.

The ability to access and analyze information about compositional and technical similarities is important.[5] For users of an imagebase, information about color, texture, and light may be critical. End users may want to search using descriptive terms like "bright red," "cityscape," "cows," and so on. Drawing interpretative relationships between and within images is presently a human task. The user's interpretation of content shares with the indexers assignment of subject terms an individual orientation and therefore, a lack of standards and consistent application of terms. This matter is further complicated by levels of representation. A clock for instance might represent time to one user and an example of a type to another.

Inconsistent application of terminology is a thorny problem for all fields concerned with classification or indexing of objects. Several recent articles have empirically studied error in archaeological classification. Fish looked at the classification of both typologies and attributes and found that personal bias can play a part.[6] Beck and Jones found that systematic error is introduced into the classification of archaeological objects in three ways:

1) Variation in the explicitness with which attributes are described and definitions are applied;
2) Differences among different analyst's perceptions; and
3) Changes in an analyst's perceptions over time. Bias can also be introduced during data collection.[7]

These same types of problems occur in assignment of index terms to document and object and image records.

There has also been a lack of concern with indexing of images in print- based materials. Few back-of-the-book indexes (with the exception of art books) contain references to graphic materials contained within the text.[8] **Preliminary research has shown that images play an important part in the search process in both print and online formats.[9] Therefore, for an information retrieval system which includes graphical materials as well as text, the adoption of standard indexing and retrieval methods will result in far less retrieval precision for images and, indeed, can hamper the information retrieval process for both images and text.**

These findings suggest that it would be profitable to explore retrieval methods for images in computerized information systems in addition to textual indexing and thesaurus construction. **Computerized analysis and retrieval of images could offer both the needed flexibility and the consistency of application required in searching an image database.**

Visual Thesauri and Scope Defined

For the purposes of this paper, we will define visual thesauri as *collections of images in a hypermedia environment*, as opposed to a collection of images in a printed visual dictionary such as *The Facts on File Visual Dictionary*. A thesaurus originally simply meant a "storehouse" or "repository, as of words or knowledge." Peter Mark Roget published a list of words in 1859 that he categorized as synonyms, antonyms, or related words. Because of the great utility of his work the word *thesaurus* has come to be associated not only with a textual storehouse but one which, by its organization, reveals pre-defined structures and relationships among concepts.

For the user, the visual thesaurus in a hypermedia environment differs from its printed counterpart, a visual dictionary, primarily in its ability to skirt hierarchical term associations, its lack of reliance on textual indexing, its independence from a particular taxonomy, and its end-user orientation. Hypermedia gives us the opportunity to reformulate the current structure of the text-based thesaurus from a unified normative taxonomy to one less dependent on predefined rules about relationships presented in a textual hierarchy. Take the example of a search for images of woman and children. An initial search returns a set of images including a few WPA photographs and a painting of the *Modonna and Child*. The user is interested in the statement made by the above named painting where a baby sits on a woman's lap. Rather then having to formulate a verbal query to match this need, the user outlines the major figure and asks for similar images in terms of subject content. With the basic visual message clarified, the searcher can use text to determine media [painting, photograph, tapestry], time period [14th century, contemporary], and cultural or social factors [mother and child, wealthy or impoverished].

We propose a hypermedia visual thesaurus that would support browsing and searching by means of direct image <-> image links, without the use of text as a preliminary or intermediate retrieval device. Such a system could also offer the user the flexibility of combining several different search mechanisms, both text- and image-based, in any order. To reach this objective, the system must be able to identify the "relevant" semantic parts of an image. To do so we need to exmain the relations between verbal and visual language, what part visual information plays in problem-solving, and what types or kinds of information are best communicated visually.

Analysis of Printed Visual Dictionaries

The relationship of graphics to text can be explored through a tool which has existed for centuries: the 'visual dictionary.' The purpose of a visual dictionary is to enable a user to find a name for an item whose name is unknown. It also enables a user to find a picture of an item whose name is known, or the name of a part of some known item (for instance, the signature of a book).

In 1672, Joh. Amos Commenius's *Visible World, or, A Picture and Nomenclatura of all the Chief Things That are in the World* was published in London. Since then many pictorial

guides, visual dictionaries, encyclopedias, or indexes on a wide variety of topics have been created. Notable are: The *Oxford-Dudun Pictorial English Dictionary* (Phelby 1981); *The Stoddart Visual Dictionary* (Corbeil, 1989); and *Peterson's Field Guide to Bird Identification* (Houghton Mifflin, 1990). For the sake of convenience we label all of these "visual dictionaries." The editors of these visual dictionaries have defined their users primarily as non-specialists, the average person: "active members of modern industrial society who need to be acquainted with a wide range of technical terms."[10]

A visual dictionary is distinguished from an online visual thesaurus in that, in print form, there is an explicit structure imposed on the information, much as a regular dictionary has structure imposed by the alphabetical ordering of terms. The structure of a visual dictionary is determined by a process which includes selection of major themes (Transportation) and further subdivision of these into categories (Transportation by Railroad) and subcategories (coach car, sleeping car, dining car). Further categorization is determined by the association of items on a page with one another, as when stone is depicted with bricks, concrete blocks, and steel, all building materials. Thus, stone in this case becomes a "building material".

Access to information takes place in one of several ways. The user can locate an item for which the name is unknown using finding aids provided such as the table of contents or thematic indexes and browsing the sections until the item is identified visually, with its accompanying terminology. Alternatively, the user can browse the dictionary without consulting finding aids. When the name of an item is known the various indexes can be used to find the picture.

Within the broad thematic sections of the dictionaries, items on a page typically exhibit two types of relationships: part-whole, and similarity. Items are similar because of similar form or function or because they belong to a particular class as determined by an existing taxonomy. The two types of relationships may appear on the same page. In many cases composite or idealized objects may be used, eliminating details while retaining enough information to present an unambiguous picture. Within thematic sections, pages of related objects are displayed. For instance, 'house furniture' can include items such as dishes, lights, curtains and curtain rods, and domestic appliances as well as tables, chairs, etc. Thematic sections also bear a relationship to one another: clothing, personal adornment, and personal articles follow each other, as do architecture, house, and house furniture (see, for instance, Corbeil, 1986).

In visual dictionaries which serve as identification aids for natural objects or the animal world, color becomes an important identifying characteristic and may be the primary initial access point, as in a book concerned with wildflower identification. In addition, other more abstract concepts, such as seasonal variation and geographical location, become important facets of the object.

However, if an item is totally unknown and the user is completely unable to place it in relation to some category such as transportation, cooking, or Finches, the only access is to browse the pictures in the dictionary. Visual dictionaries exhibit an advantage over online visual thesauri in this respect; it is relatively quick and easy to browse the pages of a visual

dictionary, even though the book itself may have a large number of pages. In addition, the relationship of an object to other objects on the same page or on nearby pages is quickly and easily grasped.

This type of browsing is difficult in an online environment. Even when a number of objects are displayed on a computer screen, rapid traversal of such displays is difficult in all but the most advanced experimental systems. The user also has more difficulty maintaining knowledge of the underlying organizational structure. In a print dictionary it is much easier to grasp the fact that images of clothing, accessories, and personal articles are in approximately the second quarter of the book and near objects associated with "home;" this spatial knowledge of intellectual structure facilitates both browsing and more structured search behavior.[11] What is lost to the user is the ease of determining that underlying structure, which becomes quickly apparent with the use of a print dictionary. Various methods (maps and webs) for reducing this disorientation have been explored, yet these remain less than satisfactory.

The print thesauri examined for this research failed in some of their stated goals; they also exhibited, in the selections and labeling of associated images, some of the problems found in indexing term assignment. One stated goal was "to assign to an illustration the role played by the written definition in a conventional dictionary"[12] yet this is only partially realized. For instance, the conventional dictionary definition of a 'shovel' is "an implement consisting of a broad blade or scoop attached to a handle, used for taking up and removing loose matter, as earth, snow, coal, etc."[13] This definition describes both the form and function of a shovel. In the visual dictionary a shovel is found in two places: with Firetending Tools; and Gardening: tools and equipment. Use is implied by the larger category and by the surrounding implements on a page, but is not specified. It is more clear in the gardening context as shovel appears with rake, hoe, etc., but in Firetending it appears with a broom, poker, log tongs, on the same page as a woodstove, with a fireplace on the opposite page. The visual dictionaries reviewed all assumed a Western orientation. While one dictionary included developing countries in its list of intended user groups, much of the information found therein remained inaccessible to these users.

With these information tools, indexing terminology has been determined by 'experts' and are generally accepted as authoritative. However, no single vocabulary can cover the range of possible intended uses or interpretations of meaning. Classification schemes based on expert knowledge cannot escape cultural influences, nor can they overcome the effects of time, or terminology variations between subject domains -- the very boundaries of which are dynamic. These visual dictionaries are not user-oriented because the reader must approach from the point of view of the expert whose view is represented. Published visual dictionaries work best when the reader conforms to the tacit assumptions made by the its authors.

Existing Visual Thesauri in a Hypermedia Environment

Ximagequery

Description

Howard Besser's Ximagequery is a system for locating images in one or more collections. The search process is facilitated by the presentation of search operators in one window, a thesaurus or authority file in another window, and the surrogate images. One can identify and click on terms and search parameters at will. The user need not type to construct a search query and can confidently work knowing that they are using acceptable terms. The linking between the search terms, the images and the catalog record incorporate a number of useful functions. For example, as one browses the surrogate images, relevant information in all the windows is highlighted and updated to keep current with the user's selection. A search is initiated using index terms to build a set of images, which are then available for browsing. The user searches the database using boolean logic or partial match strategies. Once the set is retrieved the user may browse up to twenty-four small surrogate images at one time. The surrogate images offer an alternative to the verbal description also provided. Any image can be enlarged on the screen for a closer look, and compared to another enlarged image.

Another feature of Ximagequery is the "map" which graphically represents the angles and location of visual media, in relation to some subject matter in a larger context. For example, photographs from a geological site have been indexed on a map with small camera icons indicating the location and angle of the image. By clicking on one of these camera icons one can bring up the associated image.

Ximagequery is not limited to searching one collection or domain. It has the potential to concurrently search a number of collections and retrieve images from these holdings. The quality of such a search is dependent on the linking between controlled vocabularies. Besser's intent is to have his system provide "a uniform user interface to different collections, yet allow each collection to maintain its own set of descriptive and indexing terms; the "look and feel" of searching remains the same, but the searching and descriptive terminology used changes for each collection."[14] For collections of images dealing with archeology, painting, and fashion different sets of indexing terms could be used, such as AAT and Chenhall's Nomenclature. Because Ximagequery is designed for an academic community, which is traditionally oriented towards literature, (i.e. texts) users working without a foundation in subject matter and online systems may find Ximagequery less satisfactory to use.

Analysis

Ximagequery is the marriage of a standard text-oriented online library catalog with a powerful image browsing mechanism. It allows the user to browse digital images without direct intervention by library personnel and without having to thumb through the physical images themselves. When designing this system, Besser's principal concerns were the wear

and tear on images during physical browsing, saving personnel costs related to serving the user, and creating greater flexibility during the search process by providing a more dynamic search environment. These three objectives have been met in the interactive hypermedia environment of Ximagequery.

While Ximagequery is an excellent system, it is based on several important assumptions that need to be considered. One is the distinction drawn between image and text, another is about the author's intention. There is no clear distinction between the image as a work in itself, a visual representation, or a copy of a copy [slide of a photograph, e.g.]. It seems to us that Besser is talking mostly about the media through which the user encounters the item, and not some inherent quality of the item itself. According to him, with the written text, knowledge representation can be obvious even explicit; the author's intention can be known. With images representation is interpretive and textual.

NASA Visual Thesaurus

Description

The NASA Visual Thesaurus is based on Project ICON located at the University of Texas, Austin.[15] The project deals with the automation of image management in an academic or large institutional environment. Its primary focus is the automation of image cataloging, and the image retrieval processes. Of particular interest to us is their work on the substitutability of images for textual descriptions in an image retrieval system.[16] It was this line of inquiry that lead project researchers to the idea of a "visual thesaurus" as an aid to all users of the system, catalogers and searcher's alike. The NASA Visual Thesaurus functions much like a traditional text-based thesaurus in that it provides the user with related, broader and narrower terms along with corresponding digitized images.

This system links two separate but parallel thesauri, one verbal and other visual. The coupling of these is invisible to the user who is free to search either by text or image. Both hierarchies are presented together in a single user interface designed to illustrate the relationships between visual and textual representations. The objective being to offer a 'flexible alternative' for cataloging and searching the database, one that allows structured searches based on either images or terms, or by browsing the term or visual thesauri sequentially.

What the user initially sees is a main screen with three alternatives: "programs", "facilities", and "personnel." Each category is represented by a verbal term and by an image. One also has the option of going directly to the Term Index. The thesauri are operational at the main screen selection. After making an initial selection the user proceeds by choosing images or terms to narrow or broaden their search. After entering a term or clicking an image, the system responds with related, broader and narrower terms and any corresponding digitized images. The image or term once selected becomes part of the search query applied against the full database. Searching the NASA Visual Thesaurus is interactive. The results of a search are presented as both text and image. The visual representation is said to disambiguate the term assignment. The term thesauri is the authority source.

The visual interface masks the text-oriented retrieval system which constructs data sets based on statistically derived weighted values. The Visual Thesaurus does not use boolean search logic. Its visual interface is independent of the data retrieval mechanism used to search the imagebase. It was hoped that such an interface would not be hardware and operating system specific and might be used in a variety of situations present in the NASA environment.[17]

Analysis

Like Ximagequery, the NASA Visual Thesaurus functions as an interface to a large and rapidly growing collection. Undertakings such as The Athena Project (MIT) and Besser's work at The University of California at Berkeley (Ximagequery) have demonstrated that the technology is available to create image-oriented retrieval systems. The key issue is intellectual access not technological ability to present and manipulate images. If an image carries a great deal of information for the user which is dependent on contextual and situational factors, the assumption that meaning rests in a pre-defined set of subject terms is of limited utility to control access to the contents of an imagebase. Based on this finding, Rorvig, et al determined it was useful to construct "a visual thesaurus utilizing equivalent images... to duplicate the relationships established in the linguistic thesaurus."[18] These are presented together in a single user interface thereby giving a 'flexible alternative' to use image or text.

Rorvig describes two manifestations of what he calls substitutability.[19] The first is when a set of images is already conjoined with textual descriptions. The task then becomes to match the images to the existing textual descriptions. Rorvig resolves this situation via a "term association matrix" from which hierarchies are established. The image which has the statistically best representative characteristics becomes the image shown to the viewer. The second situation is for items in a visual collection which can be described by a small number of broad terms. Here the objective is to find the "best match between 'image exemplars' and each broad class." This situation, he argues, does not respond well to statistical methods. Instead, Rorvig has used a psychometric method to match representative images to broad classes. Part of Rorvig's contribution to dealing with image-text relations in system terms is methodological -- by establishing a general set of procedures to assign terms.[20]

We agree with Rorvig and his colleagues that the difficulties of matching the available languages to the richness of images is an intellectual problem that needs to be "overcome" to advance image retrieval. To begin to get at the visual contents of things, Rorvig wants his system to enable us to search by the "language of the image."[21] Like in the indexing of textual documents, the verbal descriptors available to describe the visual content of the image are often inadequate to express meaning from the user's point of view. Rorvig observes that in the case of images, the use of thesauri to control inconsistency is not effective due to the individual responses prevalent in human reactions to visual materials. We further posit that inconsistency is the reflection of creativity and the diversity of human interests, situations, and contexts. If inconsistency is to be "overcome" system designers will need to relinquish the idea of the utility of using 'words' to index 'non-verbal' understanding.

> Conclusions drawn from experiments testing the substitutability of images for textual descriptions of images in an image retrieval system suggest that human judgments obtained by exposure to images were more robust and more quickly obtained than those derived by exposure to textual descriptions.[22]

Ximagequery and the NASA Visual Thesaurus have significant distinctions and similarities. Both are text-based. Ximagequery is oriented toward the end user and towards preservation of the physical objects. In contrast, NASA Visual Thesaurus emphasizes the processing of new items in an systematic cost effective manner. Both use images to mask standard vocabulary control techniques. These systems enable the user to employ surrogate images to construct a search, but not through image-to-image links. In these systems text functions as an intermediary between the user and the data- or image-base. Seloff points out that most users of the NASA Visual Thesaurus are initially enamored with the ability to search using surrogate images, but do not always take advantage of the image-based retrieval functions. Overtime as users become more expert, they revert to the text-based retrieval functions. This suggests that there is more to designing image-based interfaces then presenting verbal expressions or visual surrogates.

The designers of the NASA Visual Thesaurus assume that there are important semantic links between images that cannot be verbally expressed. Besser does not seem to share this assumption. However, in neither system does one search by image-to-image links. Any benefit derived by having an image present is mediated by pre-coordinated text-based associations. The ability of the system to match the user's internal text or individual unbounded interpretations with a controlled vocabulary is yet to be resolved.

LEP Event Display

Description

Another example of a visual thesaurus is the event display system being developed at the Large Electron-Positron (LEP) Collider at CERN. This system captures massive amounts of data from thousands of detectors as beams of electrons and positrons collide. Each collision, or event, is instantly analyzed by a computer to determine whether or not it is "interesting". If it is, its data is recorded, otherwise the data is discarded. Randomly selected events are also recorded as a check on the algorithms that determine which events to capture. The data from these events are used to generate images which scientists then analyze visually. These displays can show the particles involved, there paths and their energies, as well as frames of reference, such as the structure of parts of the detector.

> By pointing to an object on a computer screen, physicists can call up specific information about that object. Thus, the image on the screen acts as an index to all the information gathered by the detector for an event.[23]

The ultimate purpose of this system is to allow human beings to see particles "that are trillions of times smaller than the eye can see and that move millions of times faster than

the eye can follow."[24] The scientist, then, views the images that the computer generates from the data, looking for patterns and unexpected events.

Analysis

The LEP event display system comes closest to our idea of a visual thesaurus. The interface to the database is entirely based on images. The user views the images and, by selecting a portion of one, calls up another. It will be interesting to monitor the evolution of this system during the next few years.

Description of our Prototype

In an effort to explore the boundaries of the issues we_ve been discussing, we are developing a prototype visual thesaurus that will be image-driven to as great an extent as possible. We have chosen the domain of plant leaf identification. Similar domains to which our model could be applied deal with identification of objects or images such as artifacts, bones, fossils, pollen spores, coins, weather patterns, and the like.

Our primary interest is to reduce, or even eliminate, the role of text in the user interface. We want to create a system that allows the user to retrieve information without the use of keywords, descriptor terms, or the like. Rather, the user should be able to manipulate the image on the screen until it matches the object in hand, and thus formulate the query. The image is manipulated on the screen by using a mouse.

The first image that the user responds to is the basic leaf/leaflet shapes; i.e. narrow, elliptic, egg-shaped, oblong, etc. Once the basic shape is chosen a representative image is presented on the screen. The screen also displays an scale reference which is a human hand or portion thereof. A small image of a hand would be displayed alongside a large leaf (e.g. hickory) whereas perhaps an image of only a finger would be displayed with a very small leaflet (e.g. rose). The user is able to manipulate the attributes of the leaf, such as its size, margin, veination, lobes, and shape to best match the object-in-hand. This is accomplished through several mechanisms that have become quite common in CAD programs. The leaf image has "handles" at control points (nodes) that can be moved to adjust the shape. A control is provided to adjust the overall size, moving the control one way increased the size, the opposite way decreased it. Selecting the margin of the leaf reveals a palette of margins from which the user can choose. Selecting a vein or the petiole likewise produces palettes which allow the user to choose one that matches. The color of the leaf image is also adjustable by the user. The user can view either side of the leaf, and manipulate properties unique to the top or bottom.

As the user modifies the image to match her specimen, the system displays the number of likely matches. This allows the user to identify a highly unusual leaf with a minimum of effort by not having to adjust all aspects of the image before a single match is found. At any time the user can browse images of leaves that "match". The closest match appears in the center with the others radiating outwards in order of decreasing similarity. The user can

> Conclusions drawn from experiments testing the substitutability of images for textual descriptions of images in an image retrieval system suggest that human judgments obtained by exposure to images were more robust and more quickly obtained than those derived by exposure to textual descriptions.[22]

Ximagequery and the NASA Visual Thesaurus have significant distinctions and similarities. Both are text-based. Ximagequery is oriented toward the end user and towards preservation of the physical objects. In contrast, NASA Visual Thesaurus emphasizes the processing of new items in an systematic cost effective manner. Both use images to mask standard vocabulary control techniques. These systems enable the user to employ surrogate images to construct a search, but not through image-to-image links. In these systems text functions as an intermediary between the user and the data- or image-base. Seloff points out that most users of the NASA Visual Thesaurus are initially enamored with the ability to search using surrogate images, but do not always take advantage of the image-based retrieval functions. Overtime as users become more expert, they revert to the text-based retrieval functions. This suggests that there is more to designing image-based interfaces then presenting verbal expressions or visual surrogates.

The designers of the NASA Visual Thesaurus assume that there are important semantic links between images that cannot be verbally expressed. Besser does not seem to share this assumption. However, in neither system does one search by image-to-image links. Any benefit derived by having an image present is mediated by pre-coordinated text-based associations. The ability of the system to match the user's internal text or individual unbounded interpretations with a controlled vocabulary is yet to be resolved.

LEP Event Display

Description

Another example of a visual thesaurus is the event display system being developed at the Large Electron-Positron (LEP) Collider at CERN. This system captures massive amounts of data from thousands of detectors as beams of electrons and positrons collide. Each collision, or event, is instantly analyzed by a computer to determine whether or not it is "interesting". If it is, its data is recorded, otherwise the data is discarded. Randomly selected events are also recorded as a check on the algorithms that determine which events to capture. The data from these events are used to generate images which scientists then analyze visually. These displays can show the particles involved, there paths and their energies, as well as frames of reference, such as the structure of parts of the detector.

> By pointing to an object on a computer screen, physicists can call up specific information about that object. Thus, the image on the screen acts as an index to all the information gathered by the detector for an event.[23]

The ultimate purpose of this system is to allow human beings to see particles "that are trillions of times smaller than the eye can see and that move millions of times faster than

the eye can follow."[24] The scientist, then, views the images that the computer generates from the data, looking for patterns and unexpected events.

Analysis

The LEP event display system comes closest to our idea of a visual thesaurus. The interface to the database is entirely based on images. The user views the images and, by selecting a portion of one, calls up another. It will be interesting to monitor the evolution of this system during the next few years.

Description of our Prototype

In an effort to explore the boundaries of the issues we've been discussing, we are developing a prototype visual thesaurus that will be image-driven to as great an extent as possible. We have chosen the domain of plant leaf identification. Similar domains to which our model could be applied deal with identification of objects or images such as artifacts, bones, fossils, pollen spores, coins, weather patterns, and the like.

Our primary interest is to reduce, or even eliminate, the role of text in the user interface. We want to create a system that allows the user to retrieve information without the use of keywords, descriptor terms, or the like. Rather, the user should be able to manipulate the image on the screen until it matches the object in hand, and thus formulate the query. The image is manipulated on the screen by using a mouse.

The first image that the user responds to is the basic leaf/leaflet shapes; i.e. narrow, elliptic, egg-shaped, oblong, etc. Once the basic shape is chosen a representative image is presented on the screen. The screen also displays an scale reference which is a human hand or portion thereof. A small image of a hand would be displayed alongside a large leaf (e.g. hickory) whereas perhaps an image of only a finger would be displayed with a very small leaflet (e.g. rose). The user is able to manipulate the attributes of the leaf, such as its size, margin, veination, lobes, and shape to best match the object-in-hand. This is accomplished through several mechanisms that have become quite common in CAD programs. The leaf image has "handles" at control points (nodes) that can be moved to adjust the shape. A control is provided to adjust the overall size, moving the control one way increased the size, the opposite way decreased it. Selecting the margin of the leaf reveals a palette of margins from which the user can choose. Selecting a vein or the petiole likewise produces palettes which allow the user to choose one that matches. The color of the leaf image is also adjustable by the user. The user can view either side of the leaf, and manipulate properties unique to the top or bottom.

As the user modifies the image to match her specimen, the system displays the number of likely matches. This allows the user to identify a highly unusual leaf with a minimum of effort by not having to adjust all aspects of the image before a single match is found. At any time the user can browse images of leaves that "match". The closest match appears in the center with the others radiating outwards in order of decreasing similarity. The user can

navigate around this matrix of leaves or select any of these for further manipulation. Each leaf image also has associated with it images of a range map, fruit, and tree outline. Textual information (genus, species, etc.) can also be displayed for any leaf.

Analysis of the Problem

When considering information retrieval in general we are reminded of an old Russian proverb that goes something like: "If you know what you are looking for why are you looking, and if you do not know what you are looking for how can you find it." We are looking for alternative ways of image retrieval, ways that are less dependent on familiarity with existing taxonomies and their assigned authorities. Accomplishing this end is less clear-cut. When we began this line of research we vowed to answer existing questions and try to formulate the same questions anew -- we have done a bit of both. Based on our initial exploration of the idea of a visual thesaurus, a number of areas seem worth persuing both conceptual and technological. In this paper we have touched upon three categories related to the management of visual representations of objects. These are: human image processing, information retrieval and interpretation.and communication of meaning. These fall generally into the following categories: cognitive behavior; natural language in information retrieval; and ontology. In this section we will briefly review a few specific points, but issue a caution being that these domains overlap and are difficult to neatly package into descrete wholes.

Until recently the museum community, especially those who have traditionally dealt with paper documentation of material culture have built information systems for collections management similar to or meant to mimic traditional text-based finding aids. The online image retrieval systems largley do the same with the printed visual indexes. With this approach, knowledge or information is represented by a pre-coordinated set of terms. These systems do not allow the user (be they cataloger or patron) to address their information need or problem in a non-standard, i.e., visually oriented, way. We also found that access to visual information is similar to, if not based on, printed word dictionaries with their alphabetical arrangement; or visual dictionaries with either topical classification schemes or associations of objects such as part-whole or genus-species. These approaches we classified textual because they embody both verbal and linear knowledge representations.[25]

In visual retrieval, the important conceptual, or semantic, relationships include visual distinctions like "fatter" vs "thinner" and "shiny" vs "dull," or visual patterns that are easy to recognize but difficult to verbally express. It is our contention that these kinds of relationships do not necessarily have to be text-based, (and indeed, when they are, demonstrate the problems associated with the interpretive and contextual aspects of such values or qualities), but could be linked by visual cues and patterns and manipulated by means of graphical tools in an interactive system. In this paper, we have argued that modeling a computerized image retrieval system on traditional text-based visual dictionaries is problematic.

The shift from text-based to imaged-based retrieval brings require us to address questions. They include: how would or might people "use" images in an information system; can visual thesauri enhance retrieval when people cannot clearly articulate their information

needs; what kind of balance is needed between the richness of an image and the simplification or abstraction used as an access key; what situations are suitable for visual retrieval; what criteria do we use to make these determinations? And finally, how do people relate visual representations of objects to their textual counterparts? Interface issues, such as page and document definition and navigation, and their accompanying visual cues, are crucial to successful navigation of these systems. And finally, what is the range of associations necessary to browse objects?

If we argue for a visual thesaurus we must acknowledge some degree of linguistic relations both within and between images. What is it in an image that carries meaning for the viewer? How might we parse images so that meaningful chunks could be articulated and used to match similar patterns found in other images? Is it constructive to develop visual thesauri whose semantic structure does not mirror a printed thesaurus in a related topic? Will the semantics of "visual language" offer more flexibility to the user?

Lets consider a landscape painting as an example. How might the narrative elements in such an image be encoded in a useful, i.e. open- ended, format? This simple question belays a complex of issues, such as what carries the meaning for the viewer? Is it painter's brush technique, the pallette, the iconographic references, the situation depicted, etc, etc. Clearly these values are affected by a variety of individual and cultural factors. One approach to his issue might be to create image-based systems that allow the user to identify and search on patterns. In this context, patterns become a high level vocabulary woven togehter to form a visual language. Possible advantages are less reliance on pre-coordination of terms, although the need for a visual vocabulary control is implied. Whether one is looking at a *Modonna and Child*, at a mother holding her baby, at a baby sitting on a woman's lap, and so on, with pattern matching locatation and identification of images carring similar messages for the user is a potential solution. In doing so perhaps the user will avoid stumbling over the complextites of disambiguatation of verbal expressions. In this instance, the user makes the final determination about which image suits his or her needs. This type of matching might be done with colors, textures, and other forms of visual cues.[26]

For visual dictionaries, the only access to unnamed items is visual recognition accomplished by browsing. In contrast, an online visual thesaurus could offer more flexible search strategies for object for which the user has now words. In addition to visual recognition, users could "assemble" an image for processing or preprocessing. In order to compensate for the lack of spatio-intellectual orientation such as occurs in print dictionaries, further visual clues could be provided to the user of these systems.

Questions related to information retrieval and cognition arose during our initial exploration of visual retrieval. We raise them here to provide our readers with a sense of the complexity of image retrieval in a world dominated by text. We hope that others will join in the investigation of these matters. To date, the majority of existing hypermedia systems are precoordinated, i.e., the user follows pathways or links that have been established in advance by the system's author. If two nodes are not connected (linked) a user will not be able to jump from one to the other. Yet jumping from one "node" of understanding to another previously unrelated node is the essence of problem solving and creativity. This type of linkage

cannot be predetermined but must be generated "on-the-fly" through a cooperative effort of the user and the system. What are the limits of an interactive hypermedia system in these terms? To what extent are the problems different or the same as those for researchers in text- based systems? What level of visual complexity and what level of control over same should be presented to the user to facilitate searching?

Finally, we turn to the ontological realm. Visual dictionaries and thesauri harbor two commonly held assumptions about how people understand their experience of the physical world, how they communicate it to themselves and others. An important and significant trend in information studies is the shift from traditional search and retrieval of information objects, e.g., books and articles, to information as a value- added process.[27] That is, shifting from "information" as a noun to the verb "informing." In this context, there are two concepts of concern to us.

The first is a commonly held assumption about the distinction between an image and a text. This assumption posits texts as verbal presentations (written or spoken words) and images as non-verbal (graphic or pictoral representations or "real" or imigined worlds). This differentiation begins to crumble when viewed in non-system terms. For example, is a slide of an illuminated manuscript image or text? Are the hyroglyphics in the ancient Egyptian temples image or text?. Is an interpreation (reading) of the landscape of Vactian City image or text? It might be argued that whether what we preceive something as image or text depends on reader's persepective; that what is a text to one person maybe image to another. In most systems, the basic assumption related to providing access to pictorial or graphic images includes the tacit approval of the image - text dichotomy. What we mean by image and text has considerable repercussions for both theory and method of image retrieval. By treating visual and verbal texts as conceptual equals we have several theoretical and methodological advantages.[28] We can explore the linguistic relations between images as we would between texts. We can proceed on the assumption that image and text alike are part of the intertextual web of interpretative relations. We can spend less time trying to translate images into words, and more time thinking about how to represent visual tropes and other forms of visual language.

The second assumption, which is related to the image - text issue, is that a classification scheme can or ought to be "value-neutral?" Some people argue that meaning is culturally mediated, while the others argue for a reality fixed in time and where experience is a measure of a "natural order" of things. One manifestation of the latter viewpoint is the used of a pre-coordinated classification scheme. Information systems that are so designed transform intangible experience into tangible objects. This is the traditional and most common form of information system. For those who think that reality is a social construction, the concept of hypertext or hypermedia is appealing. The potential of limitless associations fits well with the pluralistic view of information seeking. [29] Images can, indeed should, play a central role in the dynamic information systems now being developed.

Conclusion

Those who catalog objects are just now beginning to recognize the inadequacy of verbal language as a means of recording descriptive information on material culture or the physical world in general. The use of non-verbal representations has great potential to redress the limitations of text-based taxonomies. One of the most successful approaches to library and archives automation has been through thesauri and authority control. In this paper, we have briefly described a similar process for visual media. Just as with textual descriptions, visual knowledge representation has multiple layers. The following table outlines the levels of complexity inherent in any information system dealing with visual information.

Funtions	Examples
Iconographic	A man in a suit and tie = establishment A "key" = St Peter
Sensory values	Color, texture, affective, emotional
Temporal and Spatial	Spatial relationships, perspective, change, cause and effect
Social / religious	Icons, personification, anthropomorphism
Object identification	Maker's marks, chops, styles, object characteristics, types and kinds
Visual tropes	Visual metaphors, allegory, simile, word play, word association [e.g., pitch fork to fork in the road]

Figure 1

The ability to call forth an image is not enough, not an end in itself. Additional 'functionality' is needed. What 'functions' do we want to consider? David Bearman has suggested the additional functions of, print the image, manipulate the image, view a prior or subsequent condition, and relate to data (or images) elsewhere in system.30 These functions assume one has found suitable images to manipulate, print and view. In addition to Bearman's suggestions we add: visual 'way-finding;' graphics-based information retrieval; and further clarification of the relations between graphics and text that is not systems oriented. We hope our explorations in this area will provide a basis for further research, and that many of the issues involved will be made explicit in the course of our research.

References:

1. Murr, Lawrence E. and James B. Williams. "Half-brained ideas about education: thinking and learning with both the left and right brain in a visual culture." *Leonardo,* 21 (1988): pp. 413-414.

2. Reed, S. "Visual Images," *In Cognition,* pp. 138-169.

3. Markey, Karen. *Subject Access to Visual Resources Collections: A Model for Thematic Construction of Computer Catalogs.* New York: Greenwood Press, 1986.

4. Brooks, Diane. "System-system interaction in computerized indexing of visual materials: A selected review." *Information Technology and Libraries* June 1988, pp. 112-113

5. Roberts, Helen E. "Visual resources: Reflections on the ideal network." *Bulletin of the Archives and Documentation Centers for Modern and Contemporary Art,* 2/1986 and 1/1987, pp. 27-30.

6. Fish, P. R. "Consistency in archaeological measurement and classification: a pilot study."*A merican Antiquity,* 43 (Jan. 1978): 86-9

7. Beck, C. and G. T. Jones. "Bias and archaeological classification." *American Antiquity,* 54 (Apr. 1989): 244-62

8. Liddy, Elizabeth et al. *Index Quality Study, Part I: Quantitative Description of Back-of-the-Book Indexes.* Syracuse: School of Information Studies, 1990

9. Jorgensen 1991

10. Stoddart Visual Dictionary

11. Jorgensen 1991

12. Facts on File (1986) and Stoddart

13. The American College Dictionary

14. Besser, Howard, "Visual Access to Visual Images: The UC Berkeley Image Database Project," *Library Trends* 38:4(Spring 1990):798

15. Rorvig, Mark, et al. *Automatic Image Classification by Psychometric Mapping Austin: Graduate School of Library and Information Science,* The University of Texas, Project ICON Image Scaling Laboratory, Nov 1987. [Working paper No 87-2].

Seloff, Gary. "Automated Access to the NASA-JSC Image Archives", *Library Trends,*

38:4(Spring 1990):682-96

16. Mark Rorvig, "The Substitutability of Images for Textual Description of Archival Materials in an MS-DOS Environment," in *The Application of Micro Computers in Information, Documentation and Libraries,* eds K.D. Lehmann and H. Stroll Goebel. North Holland: Elsevier, 1986.

17. Seloff, 690.

18. ibid., 687-88.

19. Rorvig, 1986

20. Rorvig, 1987

21. Rorvig, 1987, p 57. Adopted by Rorvig from a concept initially put forth by Lois Lunin.

22. Seloff, 687

23. Breuker, H. "Tracking and Imaging Elementary Particles", *Scientific American* August 1991, 63

24. ibid., 58.

25. An interesting aside pointed out by David Bearman deals with the phenomena of defining concepts in opposition. For example, clocks with hands only became "analog" following the introduction of digital clocks. Why the term analog was chosen is not clear. The same can be said of the relationship between linear and non-linear. "Non-linear " hypertext systems gave rise to linear systems. See David Bearman, "Implications of Interactive Digital Media for Visual Collections," *Visual Resources,* 5(1989):311-323.

26. Solutions to this questions may be helped along by current research in neural networks

27. Taylor, Robert, *Value-added Processes in Information Systems* Norwood: Ablex Pubs, 1986

28. This over simplification of a complex issue.is made for the purposes of this presentation.

29. There is yet another type of system in the objective - subjective debate. These focus on a narrow user community or domain. These are built upon a local knowledge base, consciously limiting variation, but covering the topic with the 'appropriate complexity.'

This type of information system has been classified as an expert system and has largely fallen into the arena of artificial intelligence.

30. Bearman, p 316