MIDS: A System for Describing Image Content Graphically for Multimedia Design

by

Sylvain Charles Morgaine

Bachelor of Arts in Computer Science and Visual Studies
Boston University
Boston, Massachusetts
1984

SUBMITTED TO THE MEDIA ARTS AND SCIENCES SECTION IN PARTIAL FULFILLMENT FOR THE REQUIREMENTS OF THE DEGREE MASTERS OF SCIENCE IN VISUAL STUDIES

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
SEPTEMBER, 1989

Signature of the author _____
Sciences
t 11, 1989

Certified by _____
Cooper
Professor of Visual Studies
rvisor

Accepted by _____
Benton
Chairman
Departmental Committee for Graduate Students

MIDS: A System for Describing Image Content Graphically for Multimedia Design

by
Sylvain Charles Morgaine

Submitted to the Media Arts and Sciences section on August 11, 1989 in partial fulfillment of the requirements for the degree of Master of Science in Visual Studies

Abstract

The first step in creating an electronic multimedia design such as an electronic book or a hypermedia document is gathering the material.
As multimedia information becomes electronically accessible, multimedia designers will need to find the material for their application interactively.
This thesis introduces the concept of *content-knowledgeable* media, detailing its creation and representation and shows how content-knowledgeable media can enhance the information gathering process through implicit, reconfigurable links to related material.
A model and a system for describing content knowledge in images is proposed as a basis for a common representation of content information across multimedia objects.

Thesis Supervisor: Muriel Cooper
Title: Professor of Visual Studies

The work reported herein was supported by NYNEX and Hewlett Packard.

ACKNOWLEDGEMENTS

I would like to thank the following people:

Muriel Cooper, my advisor, for encouraging graphic designers and computer scientists to speak the same language at the Visible Language Workshop.

Henry Lieberman for helping me focus my ideas and for much useful criticism.

Ron MacNeil and Gregory Zack for useful comments and encouragement.

Glorianna Davenport for much support throughout this thesis.

David Koons for spending the time to help organize this thesis and for sharing similar research interests.

Patrick Purcell for late night chats.

Bob Sabiston for providing a window on the C side and Russell Greenlee for closing it when the garbage collector drops by.

Suguru Ishizaki for sharing some of his design knowledge and for spending many hours with me imagining the size of objects in mental images.

David Small, Laura Robin and Ming Chen for much support and friendship.

Anne Russell and Marie Crowley for much help in correcting this thesis.

Linda Peterson for dealing with graduate students in general, and me in particular.

NYNEX for supporting me during these two years.

Finally, I would like to thank the Media Laboratory for the opportunity to work in a multi-disciplinary environment.

I have not been able to find a term that I need to denote a kind of connection or relation, approximation, closeness, allied character, between ideas. The only psychological term I know of that expresses connection between ideas is "association," but this has quite a definite meaning and one that will not do for the meaning I have in mind. The "connection" of ideas, as I call it in the absence of any other term, is quite another thing from the "association" of ideas. In making experiments on the connecting of ideas, it is necessary to eliminate the "associations," which have an accidental character not possessed by the "connections."

Selected Writings of Benjamin Lee Whorf
LANGUAGE, THOUGHT, and REALITY, 1956.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

Computers have recently enabled the integration of text, still images, full-motion video and sound as a new form of communication. *Multimedia design* involves the organization and communication of interactive *multimedia information.* It focuses on the representation and interactive design of *static media* such as text or still images as well as *dynamic media* such as video segments, animation or sound. The organization of multimedia information establishes the ways in which the user can interact with the material. This structure ideally provides the user with direct access to the information as well as browsing capabilities. Interactive multimedia applications vary in styles and purpose [Meyrowitz 88] [Hodges 89] [Akscyn 87] [Hooper 88] [Christodoulakis 86a] [Wilson 87] but all of them share the goal of communicating information using different modalities. The effectiveness of multimedia communication primarily depends on when to use another medium to enhance the comprehension and assimilation of information.

1.1 Motivation

One of the main problems in creating an electronic multimedia application is finding related information across different media. As multimedia information becomes electronically accessible, multimedia designers will need to identify relevant materials for their application.

"Future multimedia information systems will provide facilities appropriate not only for concurrent storage and retrieval of data, but also facilities for creating multimedia information interactively, extracting information from large repositories of information, comparing extracted information, transforming extracted information from one form of presentation to another, and synthesizing new information from extracted and interactively created information." [Christodoulakis 86b]

Finding a representation of multimedia information that both the user and the computer can use is fundamental in such environments because it determines how the material can be retrieved and what types of operations can be performed on each medium. (i:e text, sound, video processing, editing and analysis).

In the electronic environment, text inherently contains information that can easily cross-refer to other textual material (figure 1.1a) , whereas today images and sound have to depend on verbal or surrogate descriptions (figure 1.1b). Until vision algorithms and sound recognition systems become more sophisticated, content descriptions have to be assigned manually.

Dolphin: any of various toothed whales (family Delphinidae) with the snout more or less elongated into a beak and the neck vertebrae partially fused.
[Webster's 9th]

Dolphins use the sensitive skin on the lower jaw to investigate small objects in much the same way humans use their fingertips, but they also make use of other senses that the vast majority of land mammals do not have.
[Intercontinental 88]

Figure 1.1a Words can be automatically matched by the computer

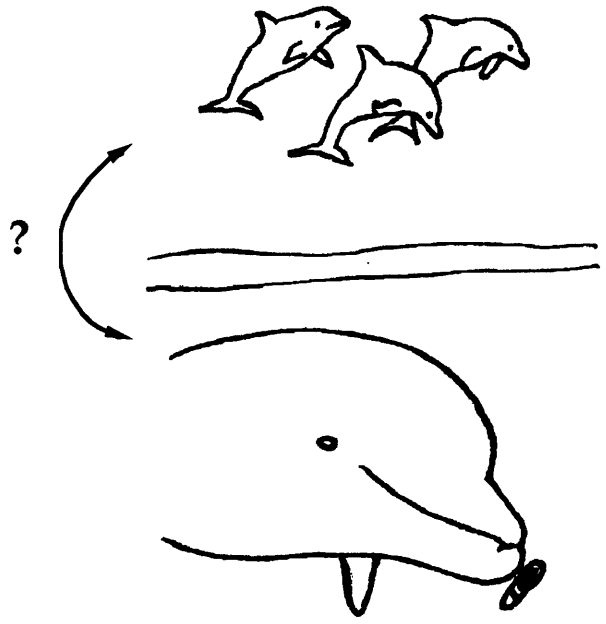


Figure 1.1b Images have to depend on verbal descriptions

Images stored electronically do not usually contain any information about their content. They remain as passive as their paper equivalent.

1.2 The need for Knowledgeable Media

In an electronic multimedia environment, images and sound need to provide the same level of interaction as their textual counterpart. Ideally, one should be able to select a visual representation of some sound (figure 1.2), or point at parts of an image (figure 1.3) and find out what instrument is playing or what the name of the part is.

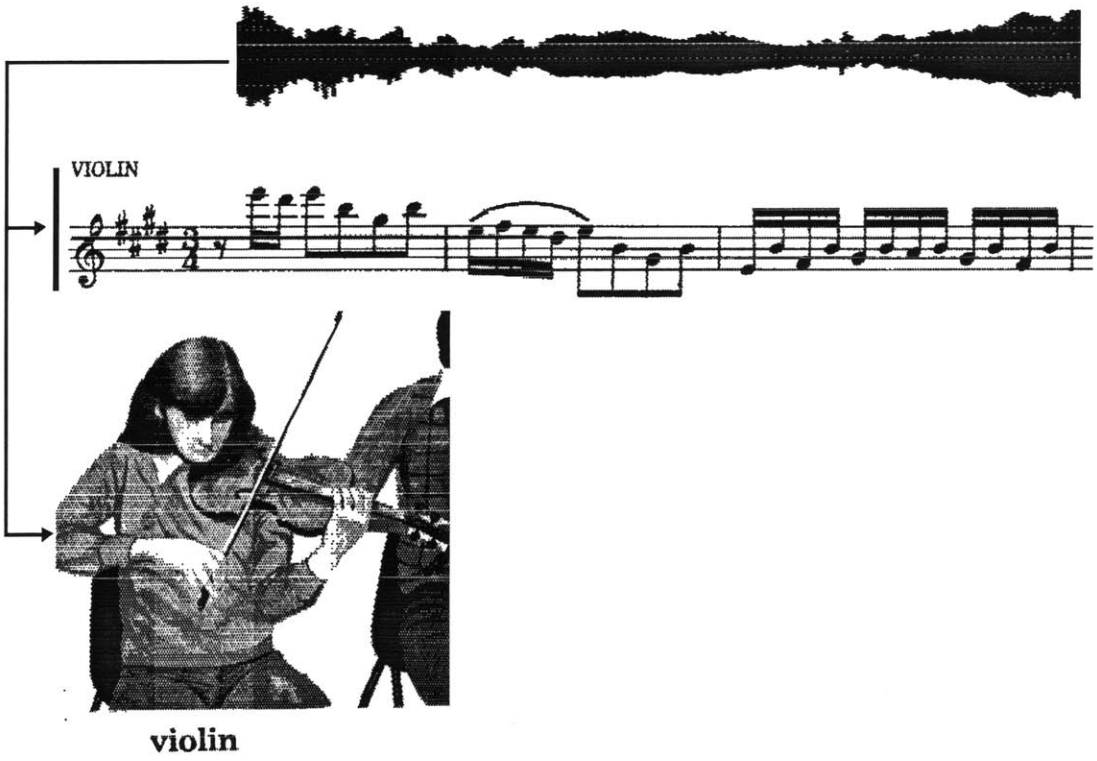


Figure 1.2 A content-knowledgeable sound sample could be used to access other representations

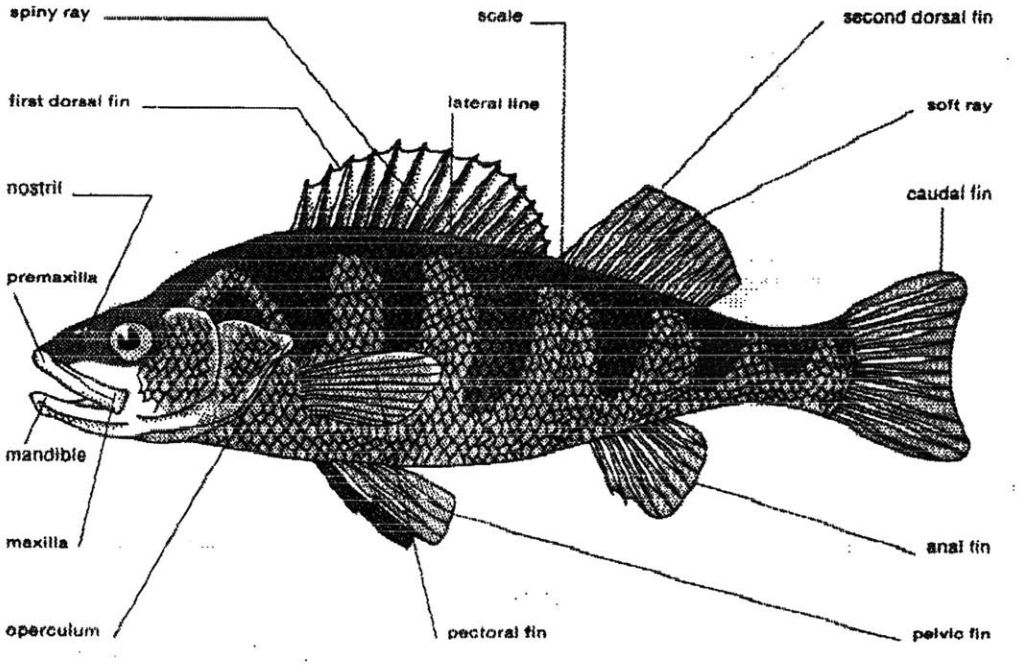


Figure 1.3 A Visual Dictionary entry

In most cases, words can be used to describe the content of non-textual media. Captions are placed under photographs to convey more information about the different elements in the image, as well as to establish the context of the picture. Narration can be used as complementary information and often brings out the more implicit details in images. Because words can be processed by machines, it then seems possible to try capture the semantic content of non-textual media into a computer representation.

Two possible methods can be used to find multimedia objects that share similar semantic content. The first one would be to store the content of each multimedia object into a database. For example, the elements contained in a video segment such as actors, actions, issues, etc. [Beauchamp 87] [Levitt 88]

[Bloch 87] [Sasnett 86] or the instruments playing a piece of music along with their individual scores. The second one is to create hypertext links that connect all multimedia objects that have similar content. The problem with both of these methods is that they do not offer, by themselves, a flexible representation for performing semantic operations which can be used to augment the information gathering process.

This thesis investigates an information pre-processing step referred to as multimedia information design. The process of information design in an electronic multimedia environment consists of gathering multimedia material, creating a semantic descriptions about the content of each multimedia object and evaluating the associations made to other material as a result of matching content relations. Once a multimedia object is knowledgeable* about its content, it becomes an active object that can communicate its information both to the machine and the designer.

A model for describing content knowledge in images is presented as a basis for a more general representation of multimedia objects. The model uses a conceptual graph [Sowa 84] to represent the information contained in an image. The graph representation is then used to associate related images.

A system called MIDS (Multimedia Information Design System) has been implemented. MIDS is an interactive system that allows an information designer[1,2] to describe the content of images graphically using a free-hand

---

* Throughout this thesis, "knowledge" means content knowledge

1 an expert, trained person, or the author of the information

2 is referred to as 'designer' throughout this thesis

sketch. The system then creates a conceptual graph representing the semantic content of the image. This representation is used to find related images.

The methodology for labeling the regions in the image is similar to creating an index for visual material in traditional books. The difference is that book indices might only have one word pointing to a whole image (figure 1.4) whereas MIDS stores a whole graph representation as the indexing reference (figure 1.5).

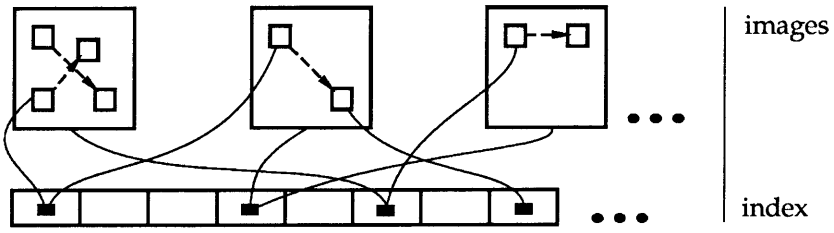


Figure 1.4 Books may only have one word pointing to a whole image

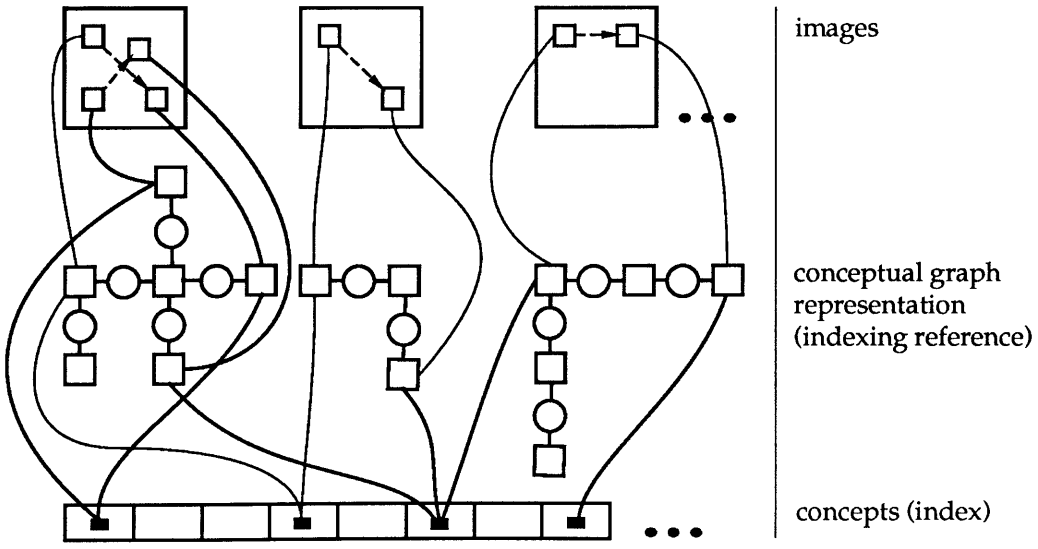


Figure 1.5 MIDS stores a whole graph as the indexing reference

The following are specific goals for this research:

- allowing the designer, who may not be a programmer, to easily represent the semantic content of images.
- providing a general representation so that it may be applied to other media.
- showing that finding related multimedia information can be made more flexible and powerful.

1.3 Related Work

MIDS was influenced by three fields of research. The first field relates to image content descriptions for picture archives. In picture archives, images are usually indexed by subject matter, author, content keywords and sometimes full-text captions [Pettersson 88]. This information structure is used to perform standard database operations but does not allow the searcher to specify actual content relations that s/he might be interested in. Because MIDS stores semantic information about the image, it not only allows one to specify content relations for the picture, but will also find related images by matching a more general classification of the elements in the image.

The second field concerns the use of link-based representations for information processing and retrieval such as semantic networks, conceptual graphs and hypertext. Hypertext systems represent information as a set of nodes and links. The nodes can be discrete units such as words or entire documents, and the links are pointers to related information. The hypertext user can either retrieve information by specifying keyword or string searches, or browse complex information networks by following the links [Conklin 87]. MIDS also uses nodes (concepts) and links (relations) to represent information, but unlike hypertext, the links are not explicitly pointing to other related information but rather are used to label the relationships between concepts within the same document, namely the image.
The analogy can be made to semantic networks [Quillian 68] [Woods 75] [Shastri 88]. A semantic network is a data structure for representing

knowledge as a network of interconnected nodes and arcs. The nodes represent concept of entities and the arcs represent conceptual relations existing between the nodes. MIDS uses a similar scheme to represent the content knowledge of the image. The difference is that semantic nets are generally used to capture accepted knowledge about particular concepts whereas MIDS is using the representation as a way of describing each picture but is not concerned with building a specific knowledge-base about the concepts occurring in the images.

Conceptual graphs are another form of knowledge representation based on linguistics [Sowa 84]. They are finite, connected graphs and generally assert a single proposition. This representation was chosen for MIDS since an image can be described as a set of individual propositions such as: "the boy is throwing the ball," "the cat is sitting on a mat," etc.

Conceptual dependencies [Shank 72] are used to represent the meaning of sentences. Shank presents three types of concepts: nominals, actions and modifiers. The nominals are discrete entities such as "elephant," "book," etc. The action is what the nominal is said to be doing and the modifier acts as an attribute for either nominals or actions. In MIDS, the nominals would correspond to concepts and the actions would correspond to relations.

The third topic of research that influenced MIDS was the manual indexing of textual material providing a more accurate content description. Salton talks about assigning special content descriptions, or profiles, to represent text content [Salton 89]. These profiles are used as a short form description of text segments and therefore can be matched faster and more accurately than using

conventional full-text search. In MIDS, these profiles correspond to each individual conceptual graph for an image.

MIDS is part of an ongoing research in the Visible Language Workshop at MIT's Media Lab on the representation, integration and manipulation of multiple media as new design elements for the designer/user. Current work includes a multimedia scripting system which begins to explore the temporal and spatial constraints of an electronic multimedia layout.

1.4 Structure of this thesis

This thesis is organized as follows:

Chapter 2 gives an overview of the Multimedia Information Design System and describes the process of representing image content graphically as well as a scenario of how the system can be used to associate images. Chapter 3 describes the implementation of the system. The creation and representation of the conceptual graph are presented in more details and the association mechanism is explained.

CHAPTER 2
OVERVIEW

This chapter introduces the Multimedia Information Design System (MIDS). The first section is an overview of the system and briefly discusses its functionality. The second section shows a typical multimedia information design session with the system. The third section presents a scenario showing how the system is used to associate images. Finally, the last section briefly describes how this representation may be extended to text, video and sound.

2.1 The Multimedia Information Design System

MIDS is an interactive graphical interface that allows the multimedia information designer to describe the content of an image graphically.

An intuitive way to assign content to images is needed as the designer may not have extensive experience with computers. Previous work at the Visible Language Workshop, MIT Media Laboratory [Greenlee 88] [Flight 89] had investigated the use of sketching to communicate concepts visually. The designer makes a free-hand sketch on top of the image to create a conceptual graph referred to as a *semantic descriptor*, representing the content of the image. The graph is comprised of two main elements: concepts and relations.

Concepts represent the regions of the image as defined by the designer and relations indicate the semantic relations between the different regions.
For example, the following image:

Figure 2.1 "A training session" (original caption)

would translate to the following graph:

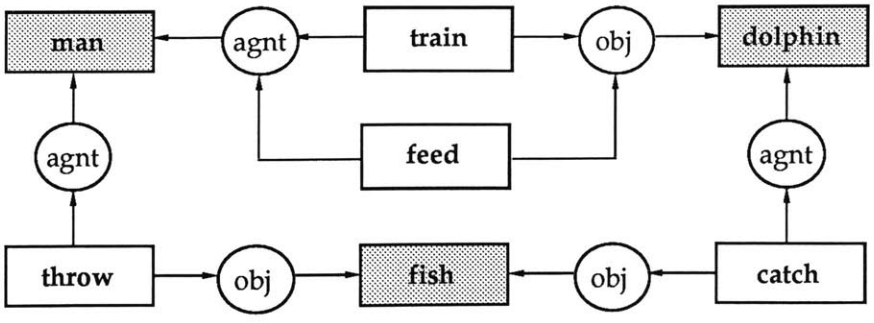


Figure 2.2 A conceptual graph representation of figure 2.1

In this graph, the shaded boxes represent the concepts in the image and the plain boxes indicate the relations between those concepts. The graph representation is explained in more details in the next chapter.

The semantic descriptor is then used to asssociate images by matching any selection of concepts and relations (figure 2.3). A detailed description of this process is given in section 2.3.

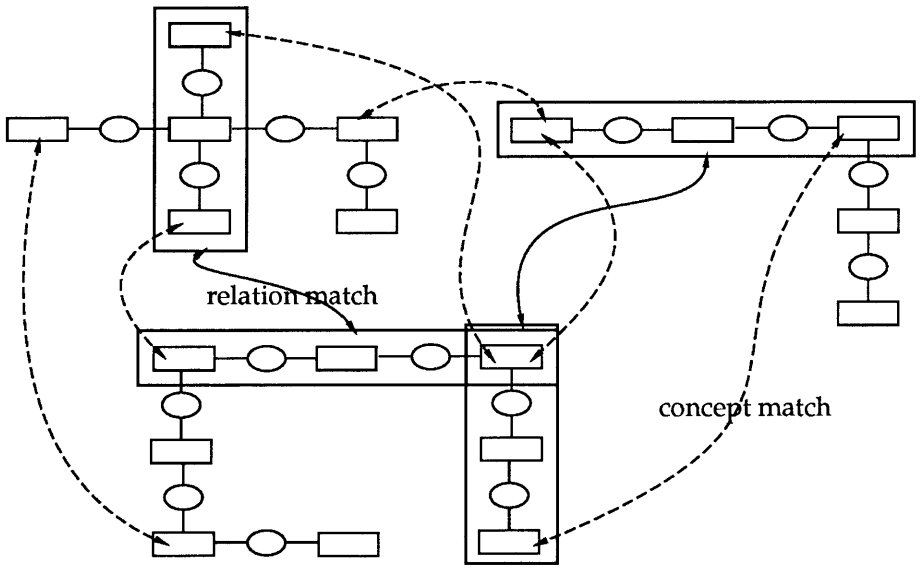


Figure 2.3 Matching concepts and relations among 3 semantic descriptors

The system is composed of two main parts: the *Graph Editor* and the *Matcher*. First the designer creates a semantic descriptor for the image by using the *Graph Editor*. The Editor is used to define concepts and semantic relations in the image. This process is accomplished by drawing directly on top of the image using the Editor's sketch functions. The image acts as a visual

template to create a semantic representation of its content. Once the image content has been recorded, the system then adds the image to the set of *Knowledgeable Images*. The Editor also allows one to delete or modify elements from the graph.

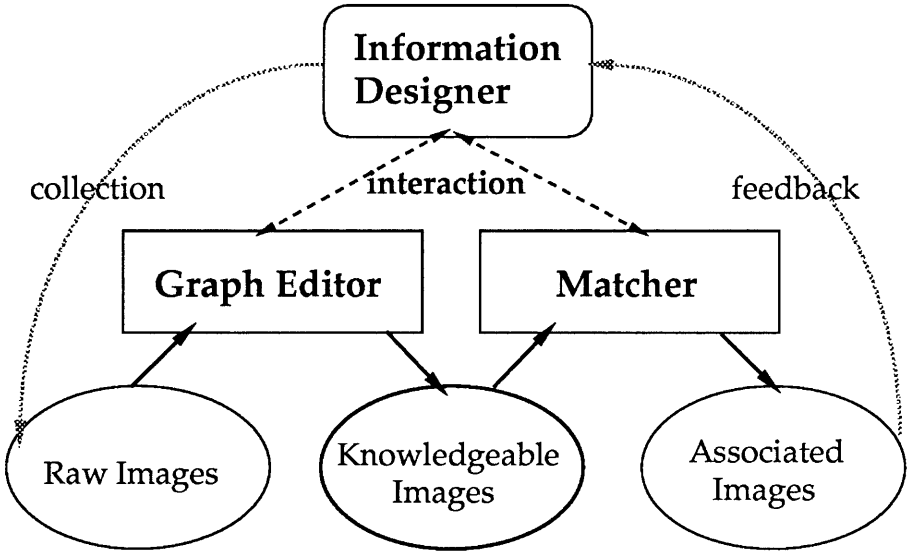


Figure 2.4 The Multimedia Information Design System

The second part of the system is the *Matcher*. The *Matcher* compares a selected image, which we refer to as the *start image*, with other knowledgeable images and returns a list of associated images having similar graphs or subgraphs.

The designer has two objectives: the first is to maximize the number of matching concepts and relations across all images by determining if any concepts defined in an image occur in other images. If a selected concept does

not produce any matches then the designer can go back and collect more images (figure 2.4). The second goal is to explore the different associations that the system can make. This process can be considered a simulation of the different types of associations that may be made by the end user when requesting new information. The designer begins with one image and combines concept and relation selections to access new information.

The system also allows the designer to encode inheritance classes for concepts as part of the image knowledge. This generalization technique automatically generates a new request to the system if no images with an exact matching of the specified concepts were found. For example, consider the previous image from figure 2.1.

Figure 2.5 The start image

Now, suppose we select from the graph the concepts *man* and *dolphin* along with the relation *train,* meaning we are interested in other images where a

man is training a dolphin. The *Matcher* first tries to find images that satisfy the exact request but fails. At this point, the system will look up the class inheritance for *dolphin* and automatically send a new request to the *Matcher* for *man* training *toothed-whales*.

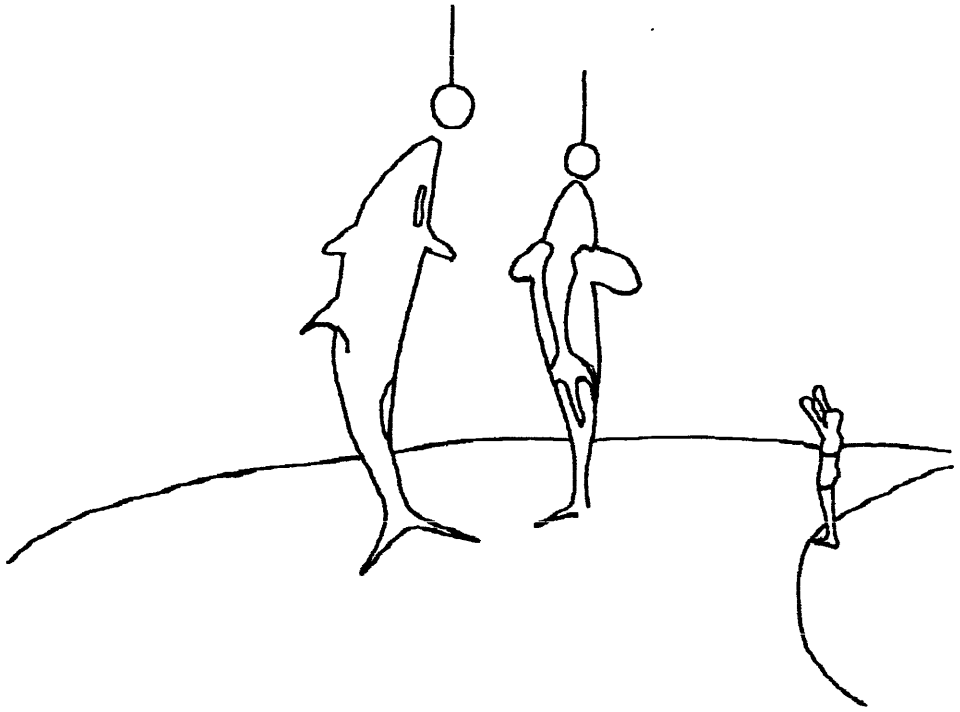


Figure 2.6 Matching a concept by generalizing

Because killer whales are also toothed whales, the image in figure 2.6 matches. In this thesis, generalizing is only applied to concepts, not to relations. The generalization process is described in more detail in the following chapter.

2.2 A Session with the System

This section describes a multimedia information design session with MIDS. The images are taken from three different sources on the subject of whales. (an educational videodisc on whales [National 81] , an encyclopedic survey on whales, dolphins and porpoises [Intercontinental 88], and a handbook of whales and dolphins [Sierra 83]). All of these sources provide factual information about the Cetacea order, specifically about the different kinds of whales and their characteristics.

The subject matter was chosen for two reasons. First, the selected images contained in the videodisc and the book are mostly close-up or medium shots showing the different features of these mammals, and have a very low visual complexity [Dondis 73]. The number of concepts in the image is very small. This traditionally provides a closer correlation between the caption (or narration) and the visuals, and for our purpose, limits the number of nodes in the graph. Second, the information in the images should be as objective as possible as the names chosen for the concepts and relations in the graph should avoid any dispute of interpretation [Salton 89]. The aforementioned sources offered accurate, factual data about whales.

In this session, the task is to create a set of knowledgeable images showing different situations between people and whales, such as whale watching or training. Ultimately, different aspects of the subject matter would be designed in a similar way to provide a multimedia designer with a wide range of

images. These images could then be browsed and selected by the multimedia designer to create his/her design.

The first step is to load the set of digitized images into MIDS. This is accomplished by clicking on **load medium** from the *data menu* (Appendix A). The system then presents a menu listing the different names[1] of images available. An image is selected from the list by clicking on its name. After selecting an image, an iconic version of the image object is displayed in the *interaction window* (Figure 2.7, Appendix A).

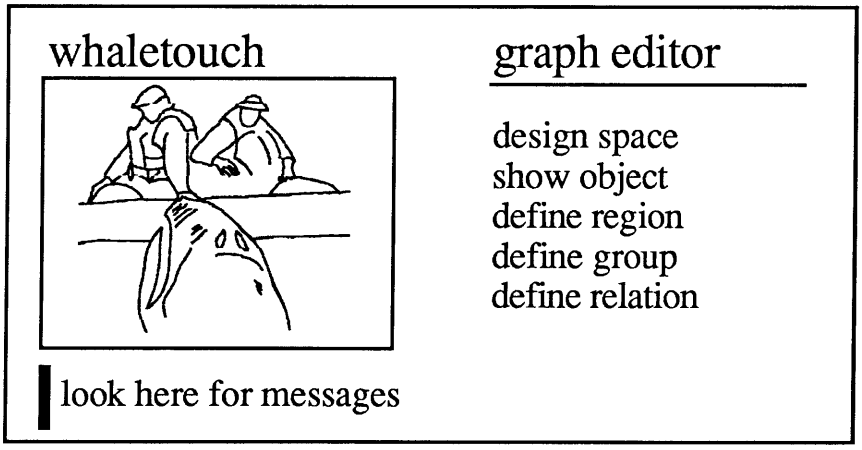


Figure 2.7 The graph editor menu

Selecting **show object** from the *graph editor menu* will display the full image on the screen. Choosing **define region** will enter sketch mode. This mode allows the designer to do a free-hand sketch on top of the image. We will

[1] The name is the actual filename for the image

show this process with one image; the other images along with their graph representation are presented in section 2.3.

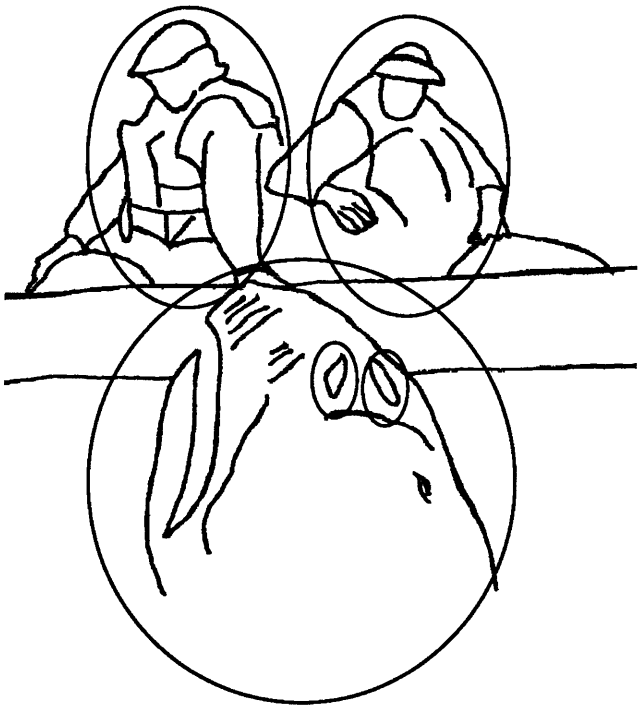


Figure 2.8 Defining the different regions in the image

The actual image (figure 2.8) shows two women on a zodiac watching a gray whale. The woman on the left is actually touching the whale.

After drawing the regions, clicking again on **define region** tells the system to create the different concepts. For each sketch stroke the system asks the designer to name the region. For example:

name that region

> **gray-whale***

Translucent rectangular regions representing the bounding box of each stroke are generated as visual cues for the designer (figure 2.9, Appendix B).

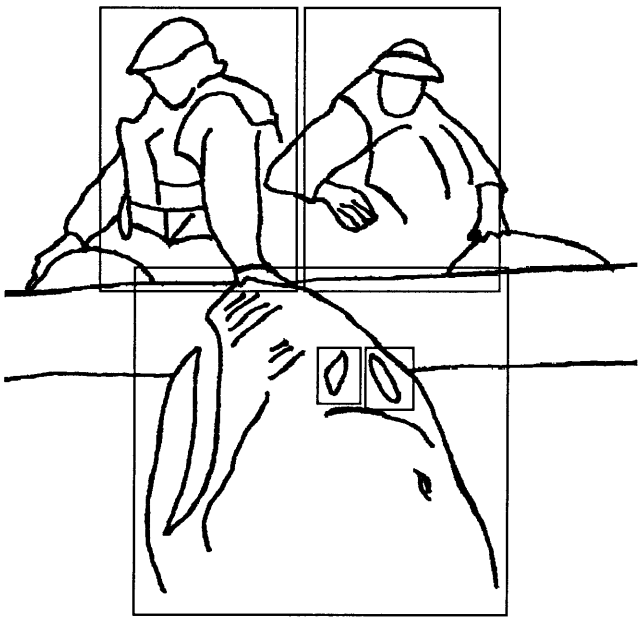


Figure 2.9 An image with regions defined

Once the regions have been defined, it is possible to draw the relations between the different concepts. Unlike the relationships used to model associations between objects in the real world, such as "Library *contains* Books" or "Satellite *travels* in Orbit" [Schlaer 88], MIDS describes specific

* Throughout this thesis, the symbol ">" means a prompt to the user and the bold type indicates the answer

relations that happen in the image, such as "The dolphin *eats* the red fish". The system supports four types of relations:

relation type	example
AGENT (AGNT) -> OBJECT (OBJ)	the **dolphin** (agnt) *eats* the red **fish** (obj)
ATTRIBUTE (ATTR)	the **red** (attr :color) fish
PART (PART)	the **flipper** is (part) of the dolphin
CLASS (ISA)	a dolphin (isa) **toothed-whale**

MIDS is only using transitive relations to represent the action between concepts. The *attribute* relation allows the designer to indicate specific properties of the concept such as color, name, weight, etc. These relations are used as additional information about the concepts. In the above example, the *part* relation informs the system that a dolphin's flipper is visible in the image as opposed to a generic flipper concept without any part-whole specification. The *isa* relation is used in Artificial Intelligence for class inheritance among concepts. In the example above: *dolphin* belongs to the class of *toothed-whales*. Concepts become more general as one moves up the chain of *isa* relations. The *Matcher* uses the *isa* relation to match on more general concepts if no images can be associated using the original concepts.
The drawing of the relations is color-coded to avoid typing in the relation type every time it is created. Figure 2.10 shows the two different color meanings.

color	concept1	concept2	relation meaning
red	yes	yes	agent -> object
red	yes	no	isa
blue	yes	yes	part
blue	yes	no	attribute

Figure 2.10 Meaning of color for drawing relations

Concept1 and concept2 are two distinct regions in the image. A "yes" under concept2 means that there is a relation between concept1 and concept2. A "no" indicates a reflexive relationship.

Choosing **define relation** from the *graph editor* menu enters sketch mode. A relation is defined as a unidirectional arrow from one region to another. A fixed but flexible syntax is used to indicate a relation between two regions. The first stroke indicates the direction of the arrow. The second stroke is the tip of the arrow and is used only as a convention but can have an arbitrary shape.

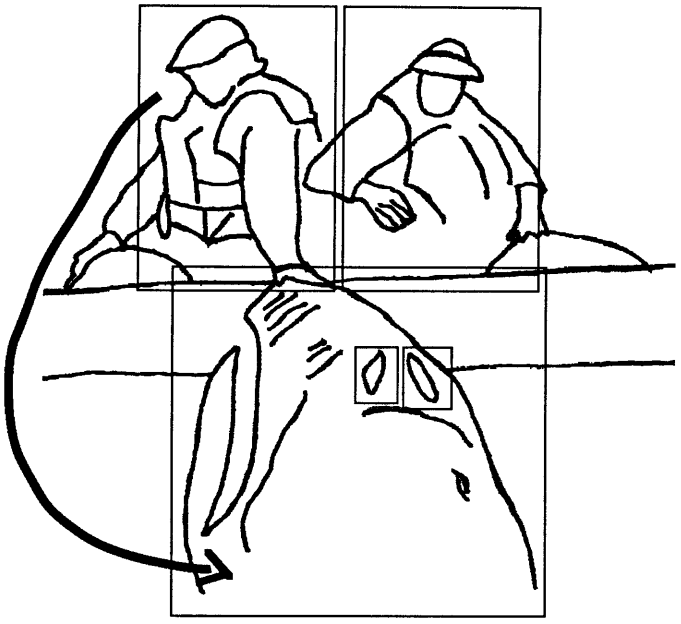


Figure 2.11 Creating a relation between two regions

The relations may be defined one at a time or as a series of strokes between the different regions of the image.

Clicking a second time on **define relation** indicates that one is finished drawing. At this time the system will parse the strokes for any syntactic errors. Basically, an arrow needs to be drawn from one region to another region, (including itself), otherwise the relation is ignored. For every syntactically correct relation, the system asks the designer to enter the name of the relation using the following format:

how is this *region-name* related to this *region-name*?

for example, in figure 2.11 the system asks:

how is this *woman** related to this *gray whale*?

> **touch**

When a relation is drawn from one region to itself, the system checks the color of the sketch. If the color is red than MIDS recognizes an *isa* relation. If it is blue than the sketch is understood as an *attribute* relation.

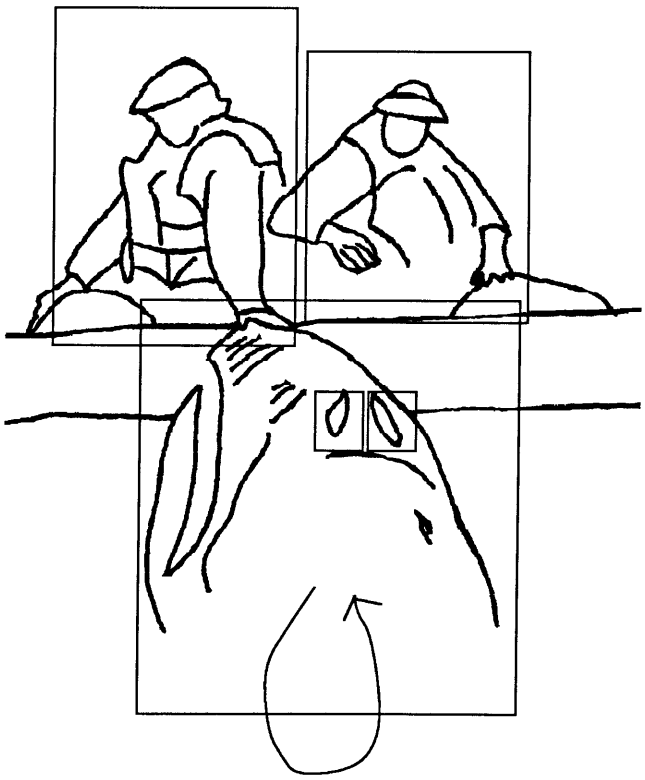


Figure 2.12 Creating an isa relation

MIDS then asks the designer to enter the name of the class:

* The italic type represents the name of the concept

a *gray whale* is a ...?

> **baleen-whale**

The next time an *isa* relation is defined for *gray-whale,* the designer can either create the class for *baleen-whale* or redefine the class for *gray-whale.* Because the *gray-whale* class was correctly defined, the first case is presented.

isn't a *gray whale* a *baleen whale* anymore?

> **yes**

ok, a *baleen whale* is a ...?

> **whale**

Following are some *isa* hierarchies as defined by the designer:

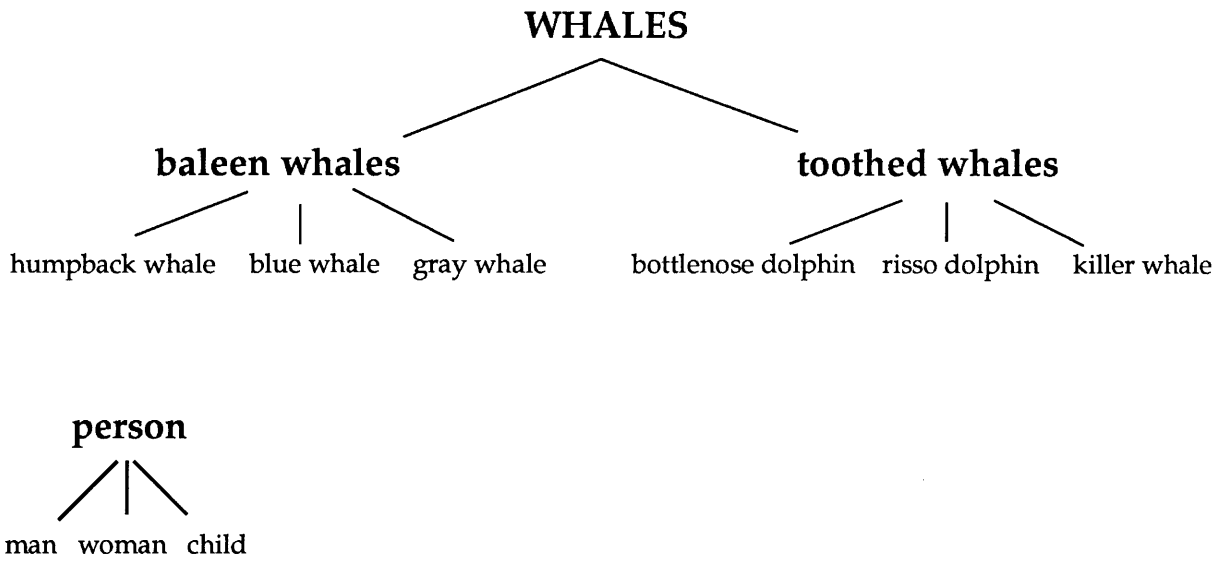


Figure 2.13 Isa hierarchies

We have shown the process of encoding content knowledge; now we shall see how MIDS can be used to browse content-knowledgeable images.

2.3 Using the Graph to Associate Images

In this hypothetical example, we will show how a multimedia designer can use the graph representation to find images. The previously defined concepts and relations now allow the designer to explore the images necessary to illustrate his/her subject. In this scenario, the task is to gather images that show contact between people and whales. We will use this goal as the basis of the interaction, although we will also show how the system allows the designer to branch off and collect other images along the way to enhance the design process. Throughout this section, we will adopt the perspective of the multimedia designer. Let us start with the image of the two women watching the gray whale from the previous section.

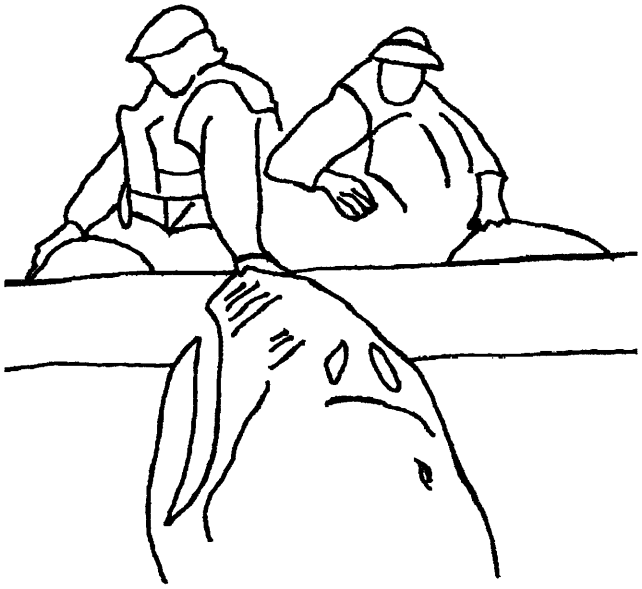


Figure 2.14 The start image

Since watching whales is a common type of contact, we will start by finding other images of people watching whales. Because there could potentially be a large number of images, MIDS allows the multimedia designer to specify how many images should be returned by the Matcher. The Matcher will generalize concepts when there are fewer exact matches than the number of images requested.

There are three concepts in the image: *woman, gray whale, blowhole;* and two relations: *touch* and *watch*. First we have to select the concepts. **Set concept** in the *matcher menu* is used to select the concepts of interest.

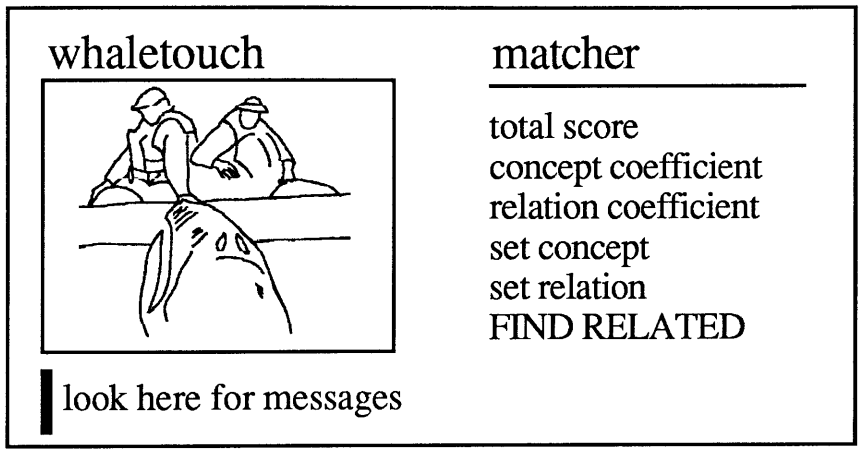
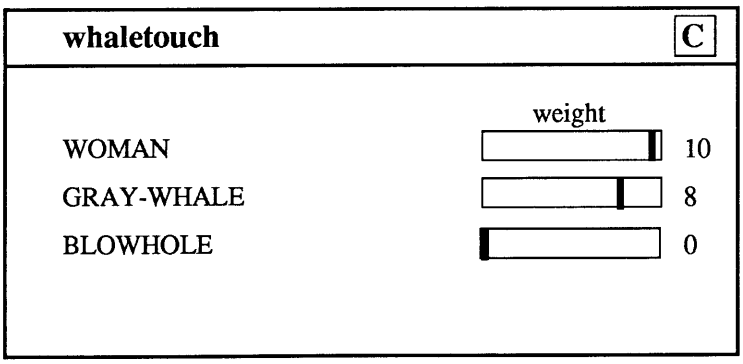


Figure 2.15 The matcher menu

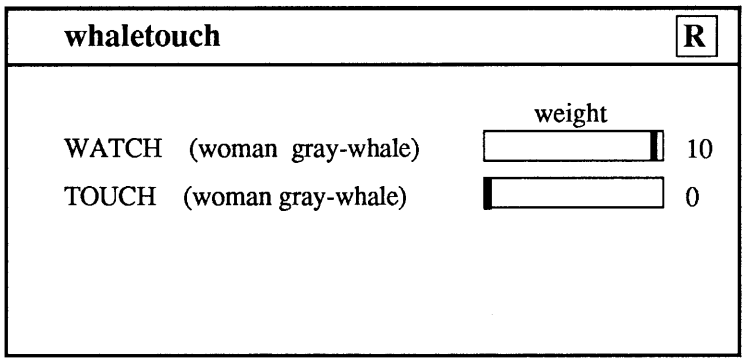
A slider menu appears showing all the concepts defined in the image. The "C" in the upper right corner stands for "concepts".



Moving the slider changes the relative weight assigned to each concept. This value indicates the user's interest in the particular concept and is used by the Matcher to determine the order in which the concepts will be generalized. For example, if we set the slider for *woman* all the way to the right, we mean that *woman* is the most important concept to match. If we then set *gray-*

whale in a lower position, the Matcher will generalize *gray-whale* to *baleen-whale** before it will generalize *woman* to *person*. This ensures that exact matches on the highest weighted concept will remain as long as possible. Since we are not concerned with looking for detailed features such as the blowholes at this point, the *blowhole* slider is set to 0.

The next step is to select the *watch* relation. Clicking on **set relation** will bring a similar slider menu showing all the relations defined in the image.



We indicate a high score for *watch* and set the *touch* concept slider to 0. Now, we can click on **find related** to initiate the association mechanism. MIDS is ultimately geared towards finding related material in multiple media, therefore the system allows the user to specify how many items per media should be gathered. MIDS asks how many images are needed. Since there are no other exact matches of *woman, watch* and *gray-whale* in our set, any number will force the Matcher to generalize.

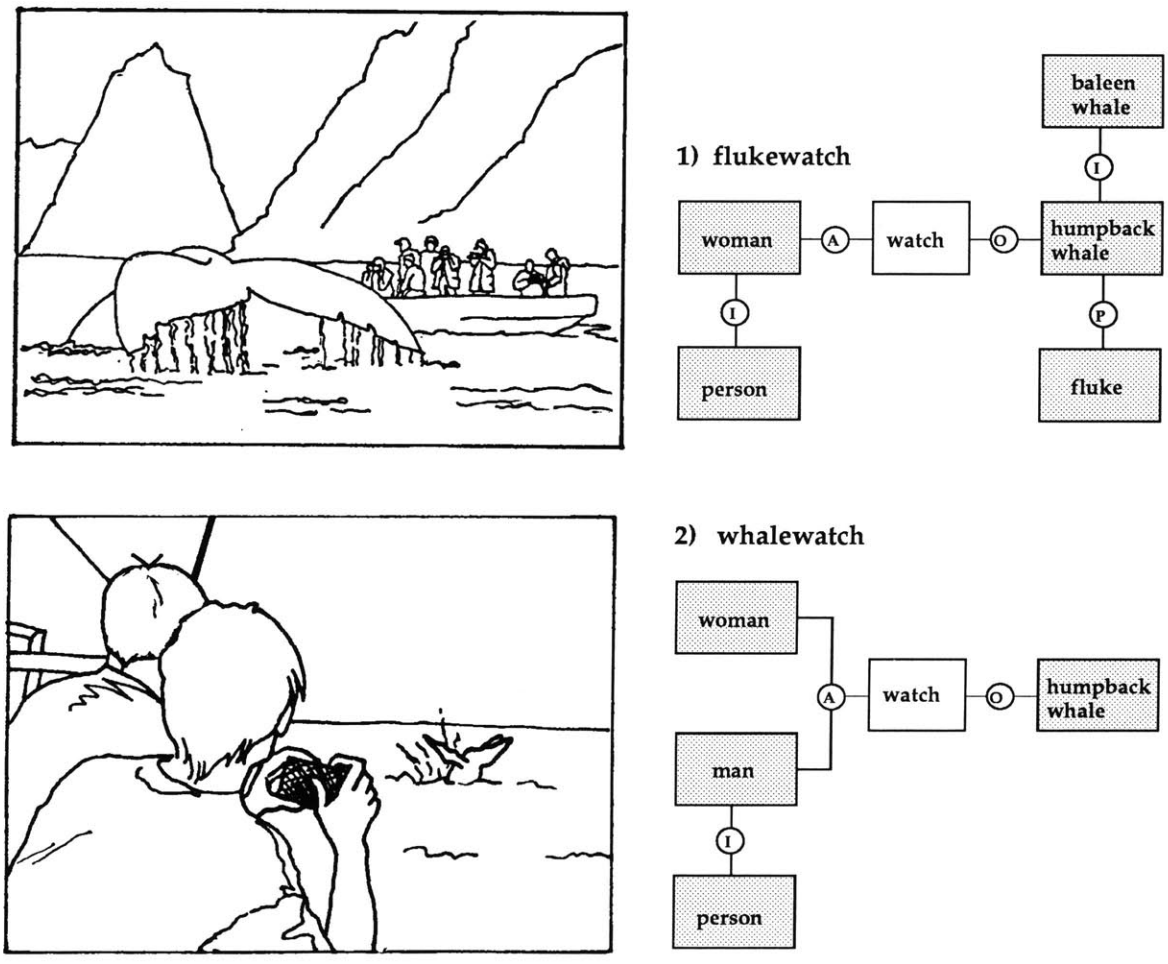
* The underline is used to represent generalized concepts

how many images?

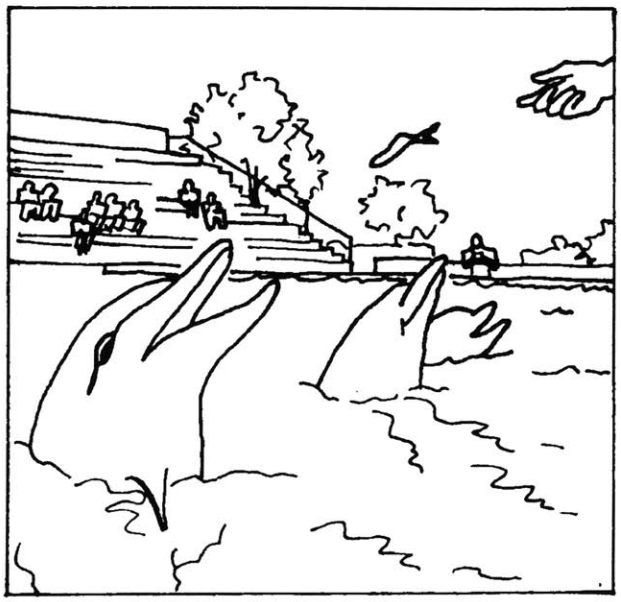
> 7

The matching function now goes through the set of knowledgeable images and returns the following*:

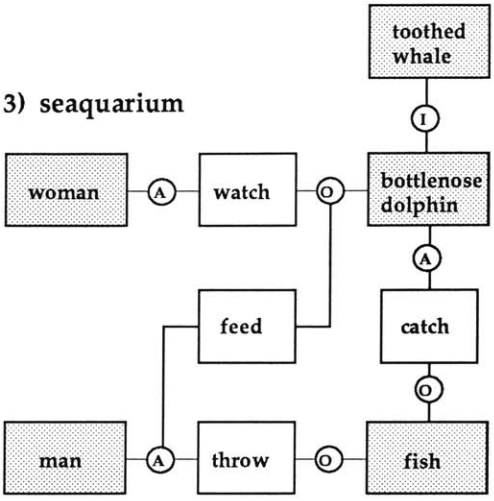
Figure 2.16 People watching whales

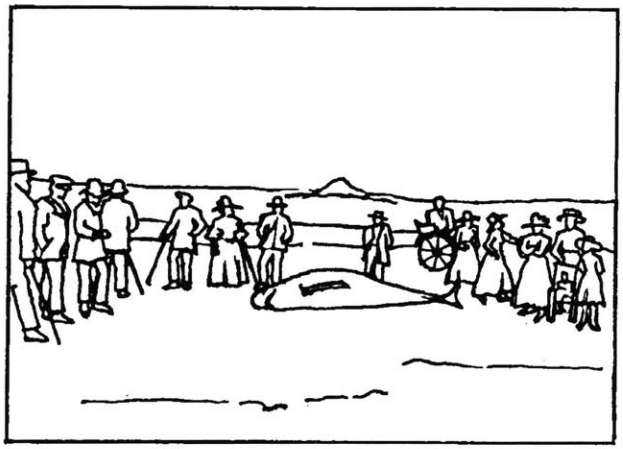


* The graphs are not actually shown in the interface but are used here to explain how the images are matched.

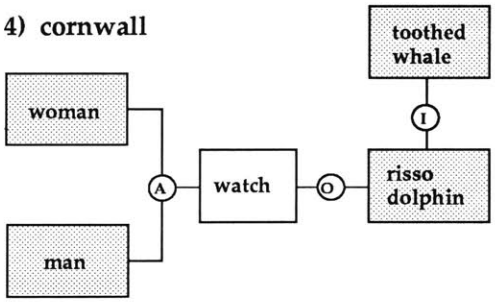


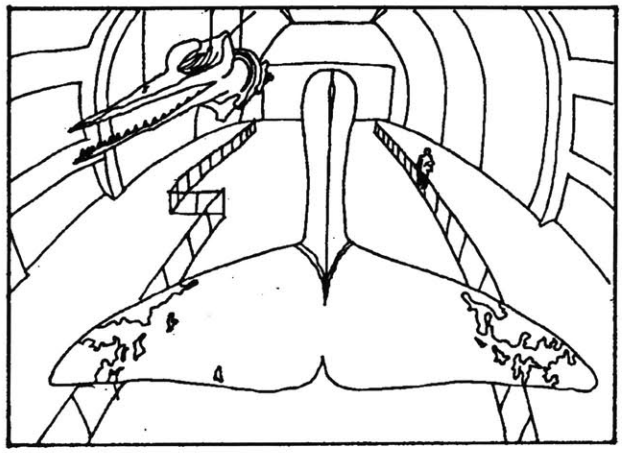
3) seaquarium



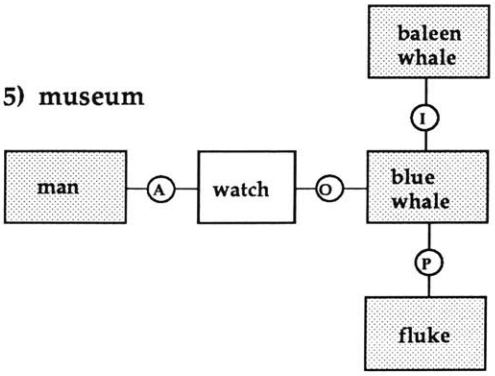


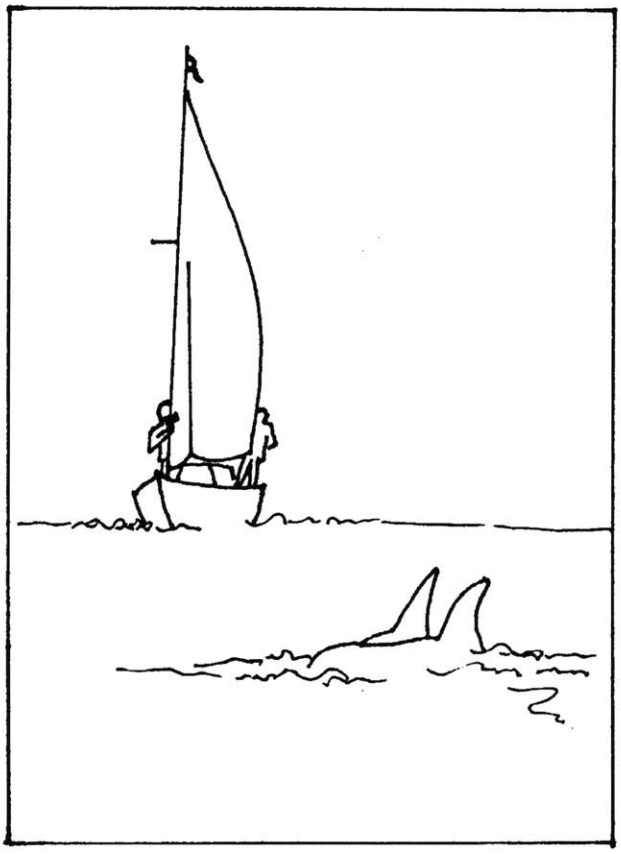
4) cornwall



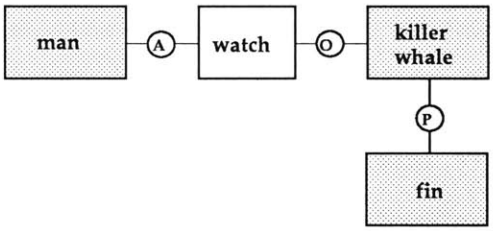


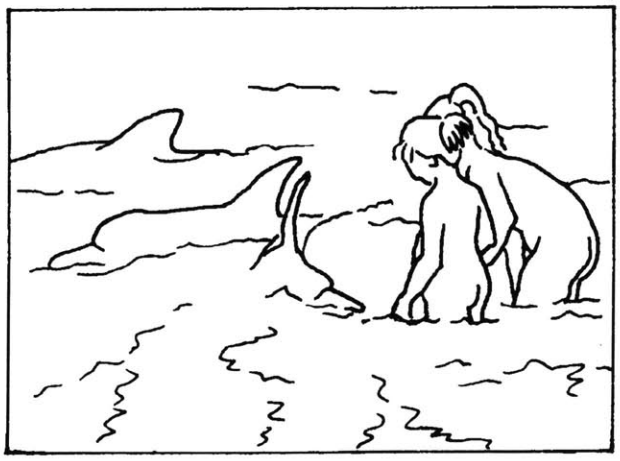
5) museum



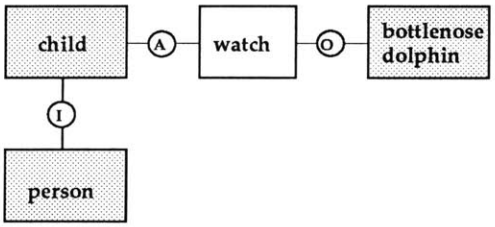


6) killerwhale





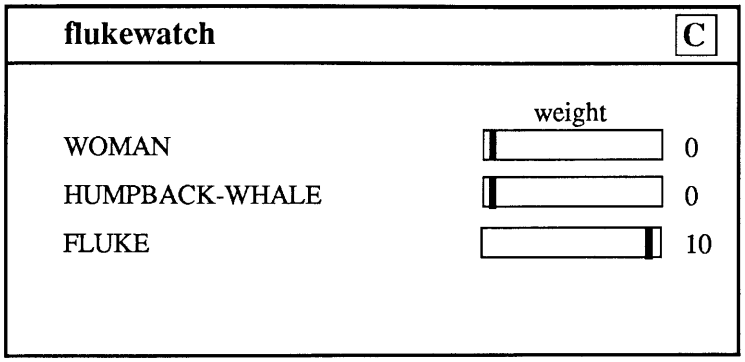
7) childwatch

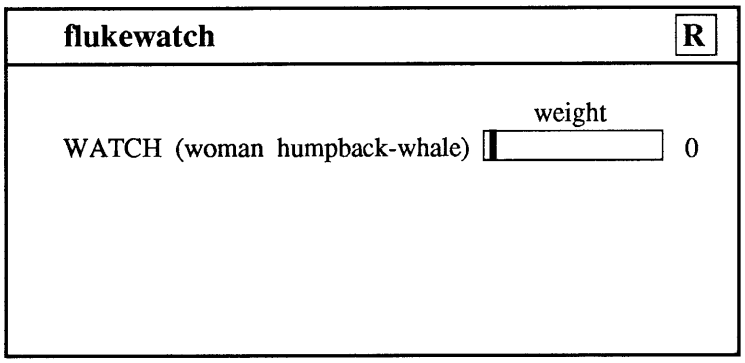


The system performed an exact match on *woman* and *watch* but failed to find *gray-whale*, as the object part of the relation, in any of the images. The Matcher then generalized *gray-whale* to *baleen-whale* and internally generated a new request for another *baleen-whale* concepts: *humpback-whale*. As a result, graph #1 and #2 are matched.

Two images out of seven are found so far. The Matcher continues its task by climbing up the inheritance hierarchy for *baleen-whale*. *Baleen-whale* is generalized to *whale*. Since all baleen whale concepts have already been explored, two *toothed-whale* concepts: *bottlenose-dolphin* and *risso-dolphin* are matched to give images #3 and #4. Now, the system explored all the *whale* concepts but did not satisfy the image count. The Matcher then generalized *woman* to *person* , and then matched images #5, #6 and #7 because the *man* and *child* *isa* relations are also *person*. A more detailed description of the match function and generalization mechanism is given in the next chapter.

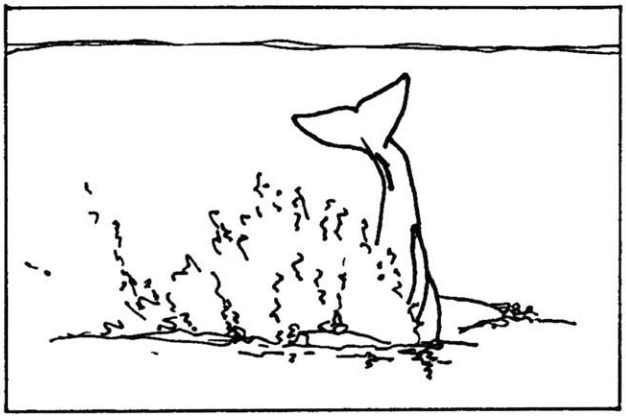
In many whale-watching activities, one often sees only part of the whale at a time such as flukes when the whale is diving, flippers, fins, etc. Supposing that we would also like to communicate this idea in our design, we might decide to gather images showing the main features of whales. We can use image #1, for example, to "sidetrack" and start another search for more images which specifically show a whale's flukes. We set the concept and relation sliders for image #1 as shown on the next page.



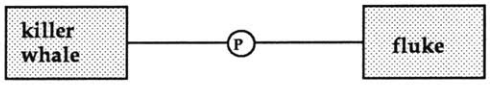


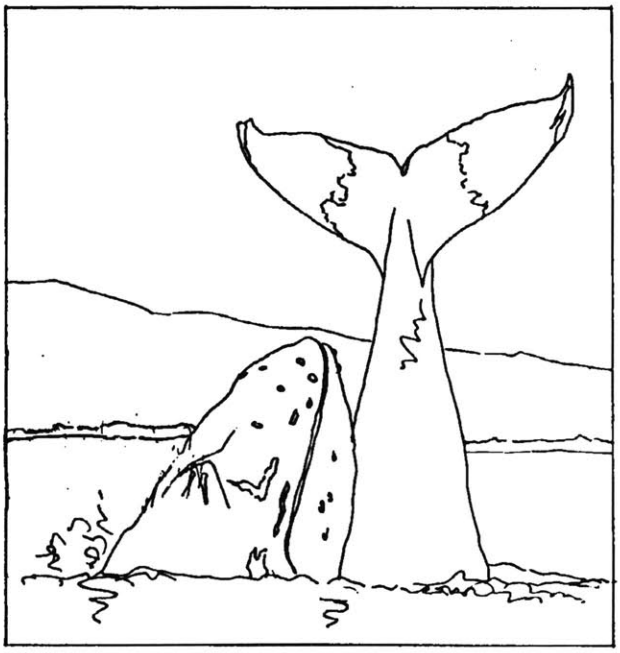
and request 5 new images. Note that since we have set the weight for *humpback-whale* to 0, MIDS will find any image showing flukes, not only flukes of humpback whales:

2.17 Flukes

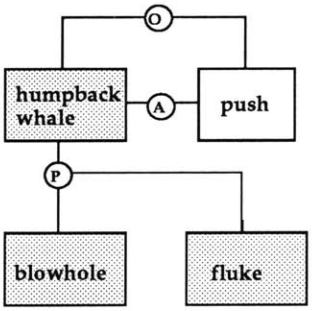


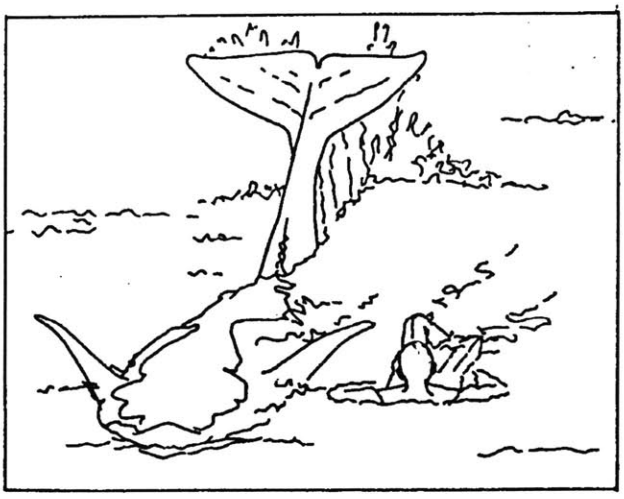
8) flukeslap



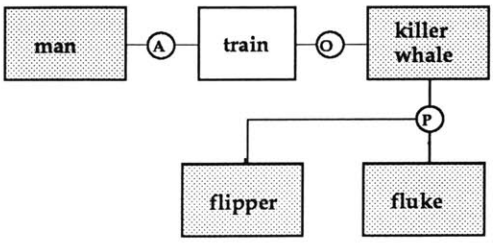


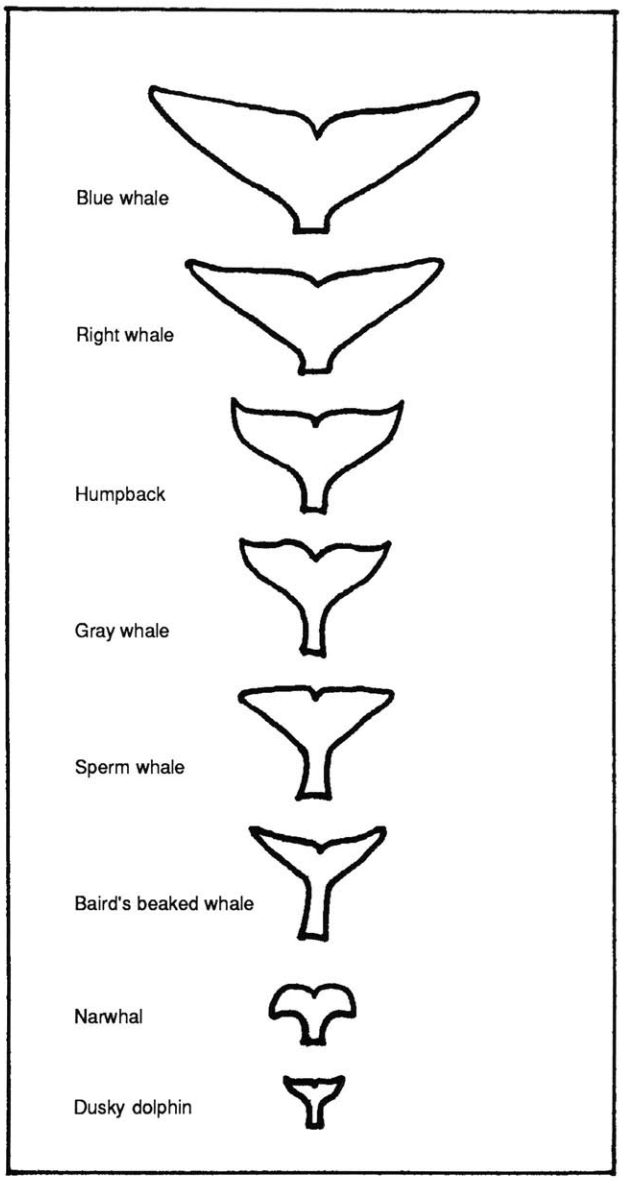
9) pushwhale



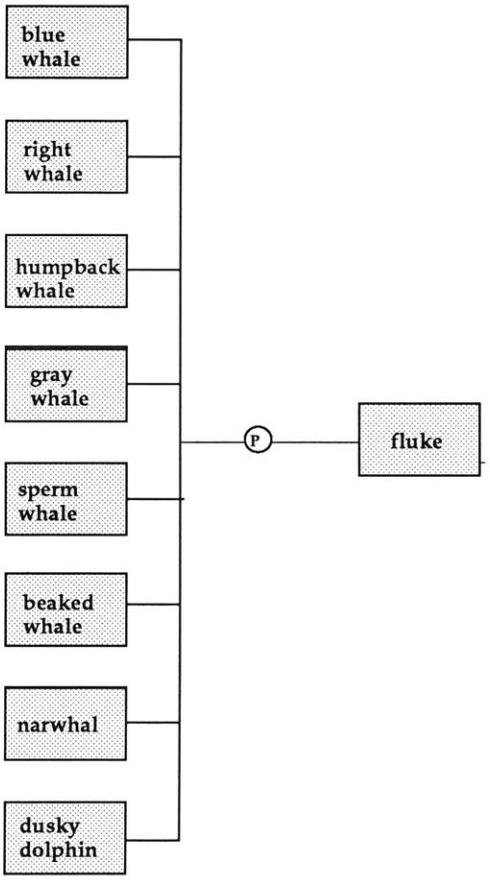


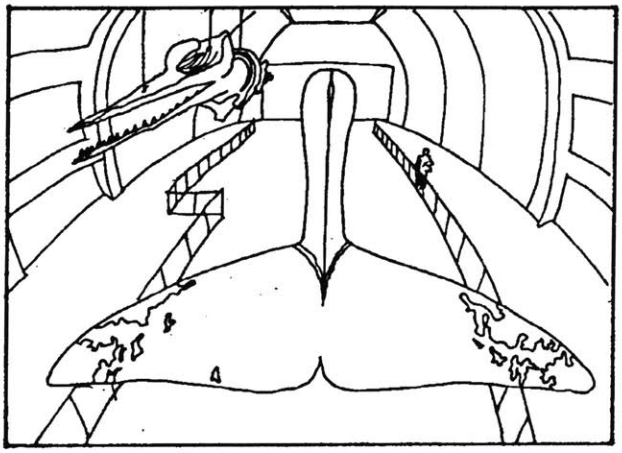
10) backswim



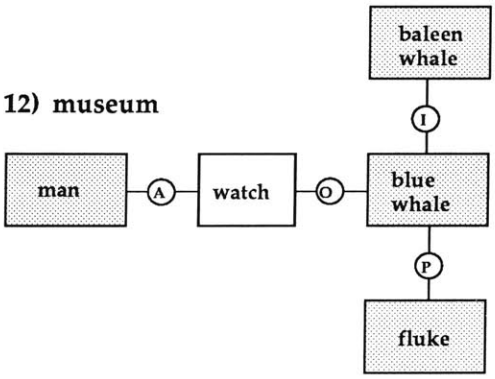


11) allflukes



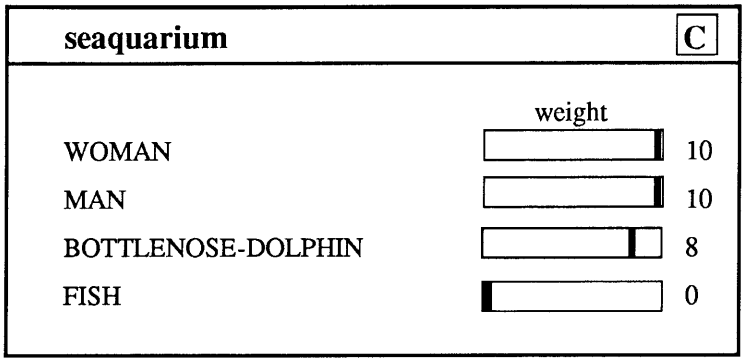


12) museum

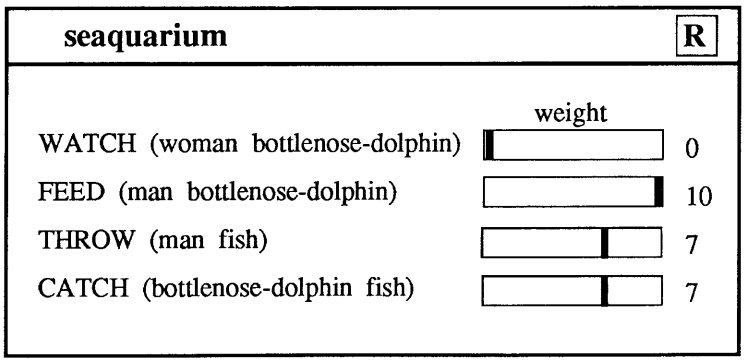


At this point we could continue our associative browsing, starting from one of the new images; however, we will go back and select image #3 to gather images of another form of contact: people feeding dolphins.

We set the concepts:

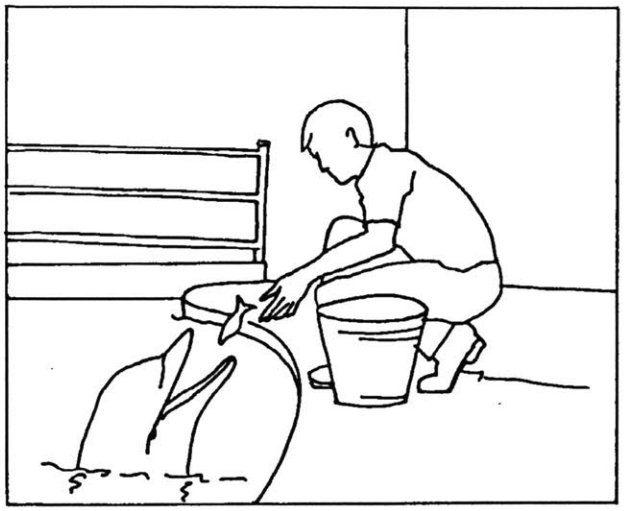


Then we set the "feed" relation:

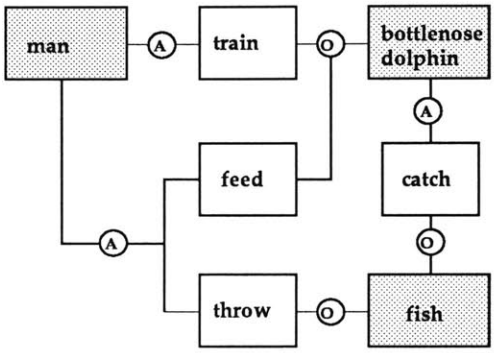


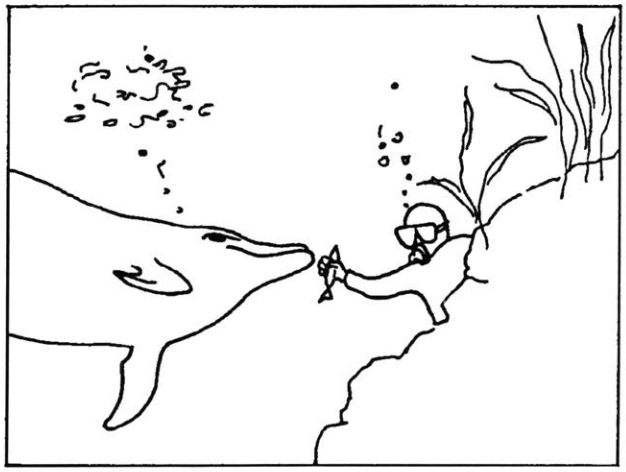
Note that we have also set the *throw* and *catch* sliders to 7, since those relations may also suggest related images. We now ask for 10 images. The same association mechanism is applied and the following images are returned:

2.18 People feeding dolphins

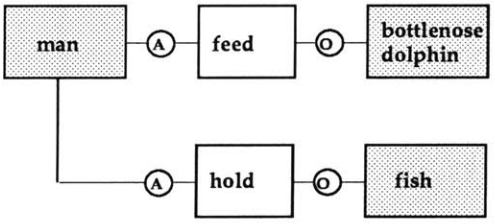


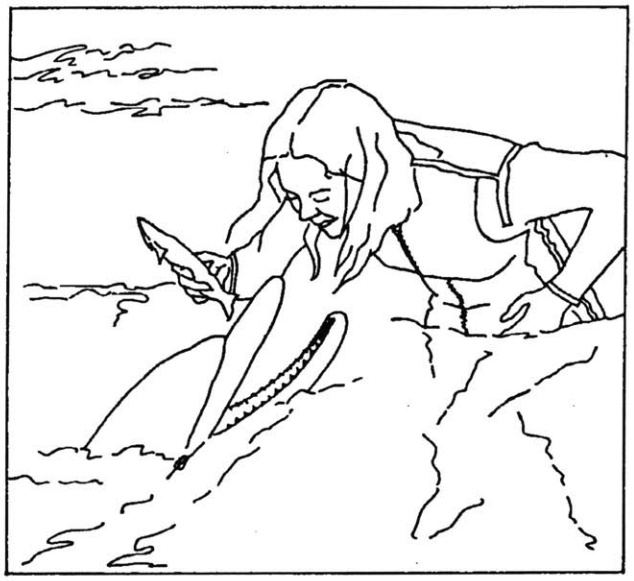
13) dolphintrainer



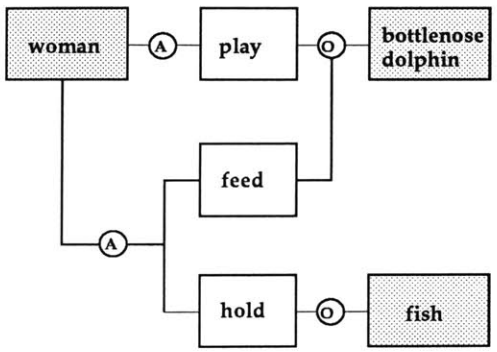


14) diver

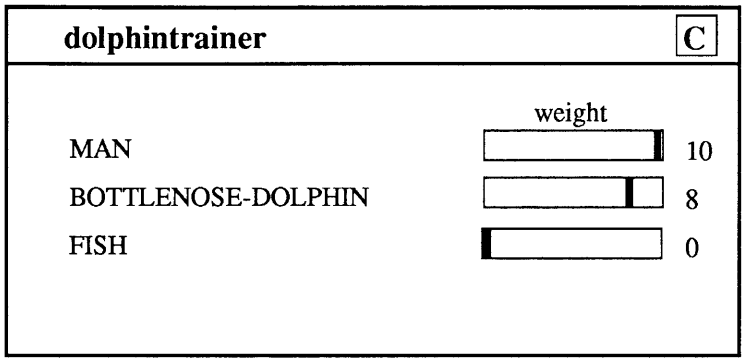


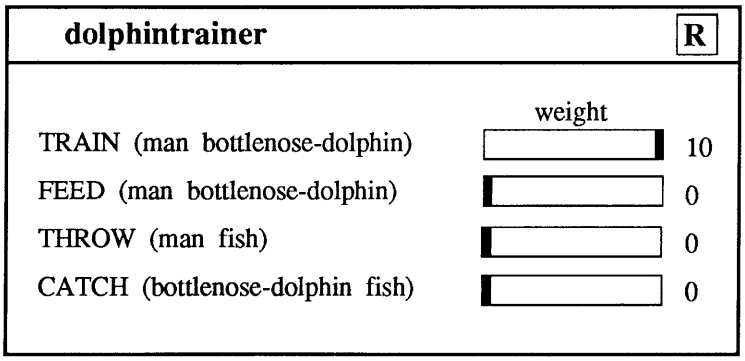


15) dolphinplayer



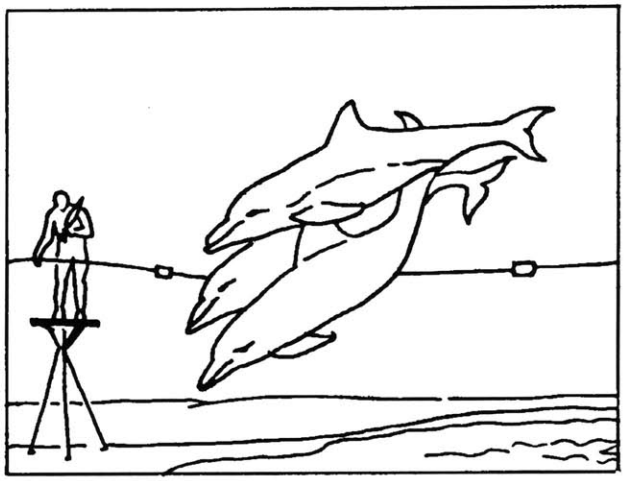
Since the system did not return more than 3 images, even after having generalized all possible concepts, we assume that no other image satisfies the request. Instead, we decide, as our final step, to explore images of people training dolphins. Image #13 will therefore be our start image:



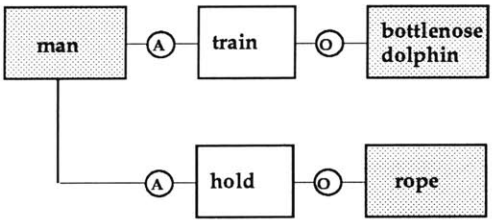


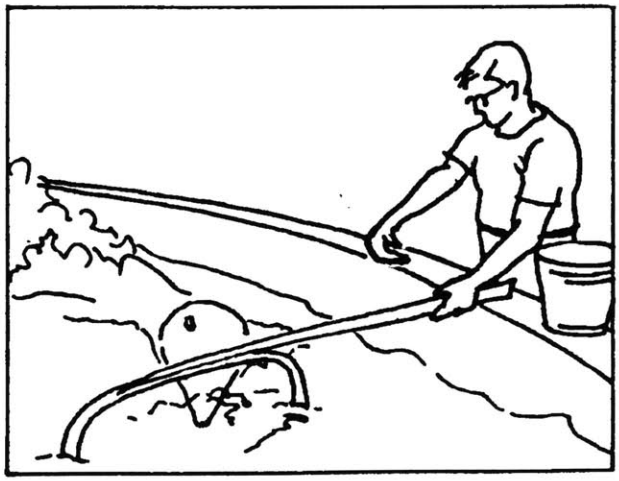
The following images are returned:

2.19 People training dolphins

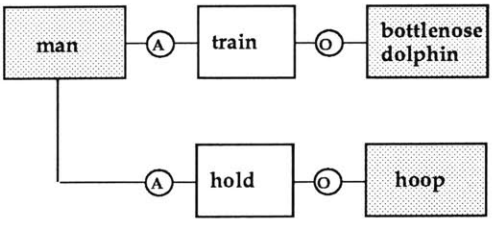


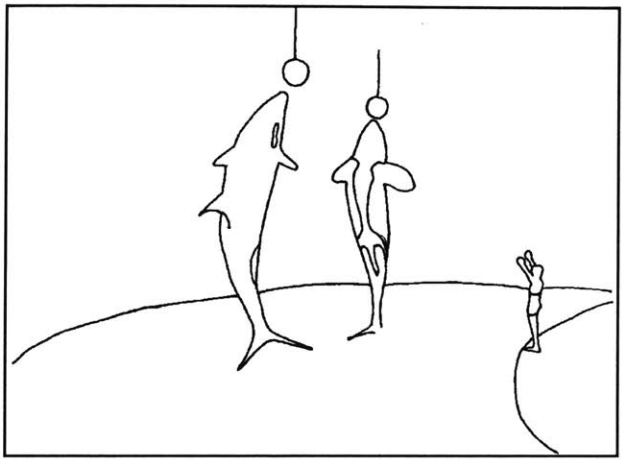
16) dolphinleap



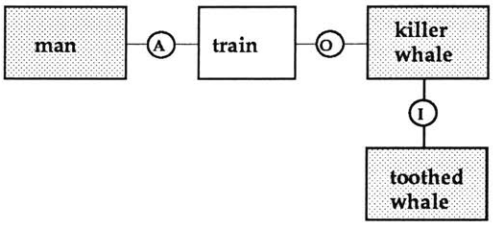


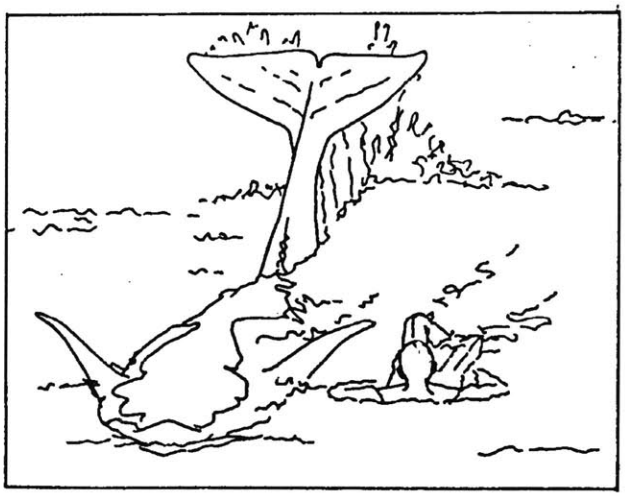
17) dolphinhoop



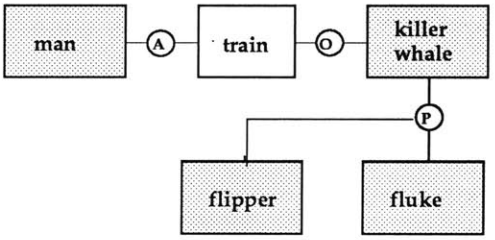


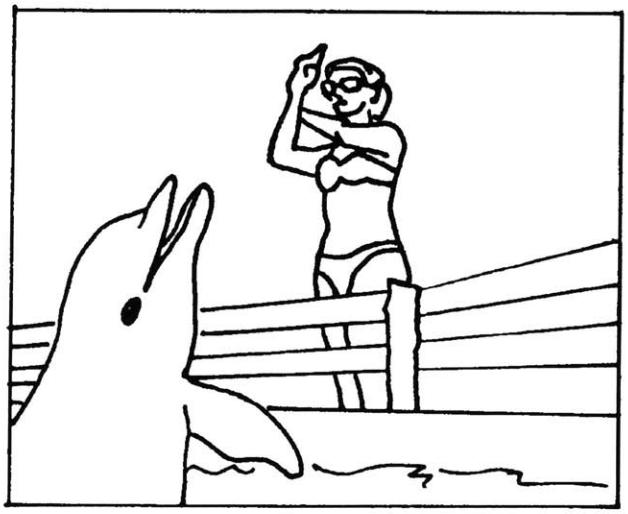
18) whaleleap



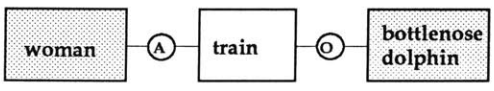


19) backswim





20) signtrainer



Note that since the information on the number of concept generalizations is always available, MIDS could graphically emphasize images with exact number of matches. However, image #6 may be as satisfactory for the designer as image #1 eventhough the former required one more generalization step. Hence there seems to be no need to impose a subjective labelling of the results.

2.4 Applying the Representation to Other Media

This section briefly describes how concepts and relations might be matched across different media.

Moving from concept to concept may not be very rewarding when dealing exclusively with stills, although this could bring unexpected images. However, concept and relation matching across different media sharing similar conceptual graph representations can be very powerful.

Creating semantic descriptors for dynamic media raises many problems. First of all, accurately describing the different types of changes which occur in filmic or sound events is very difficult. Another primary concern is that concepts and relations now change over time. If the frequency in which the changes occcur is very high then the number of graphs needed to represent these changes becomes prohibitive for the information designer. However, since the graphs can represent dynamic events, they would offer many roads for analysis.

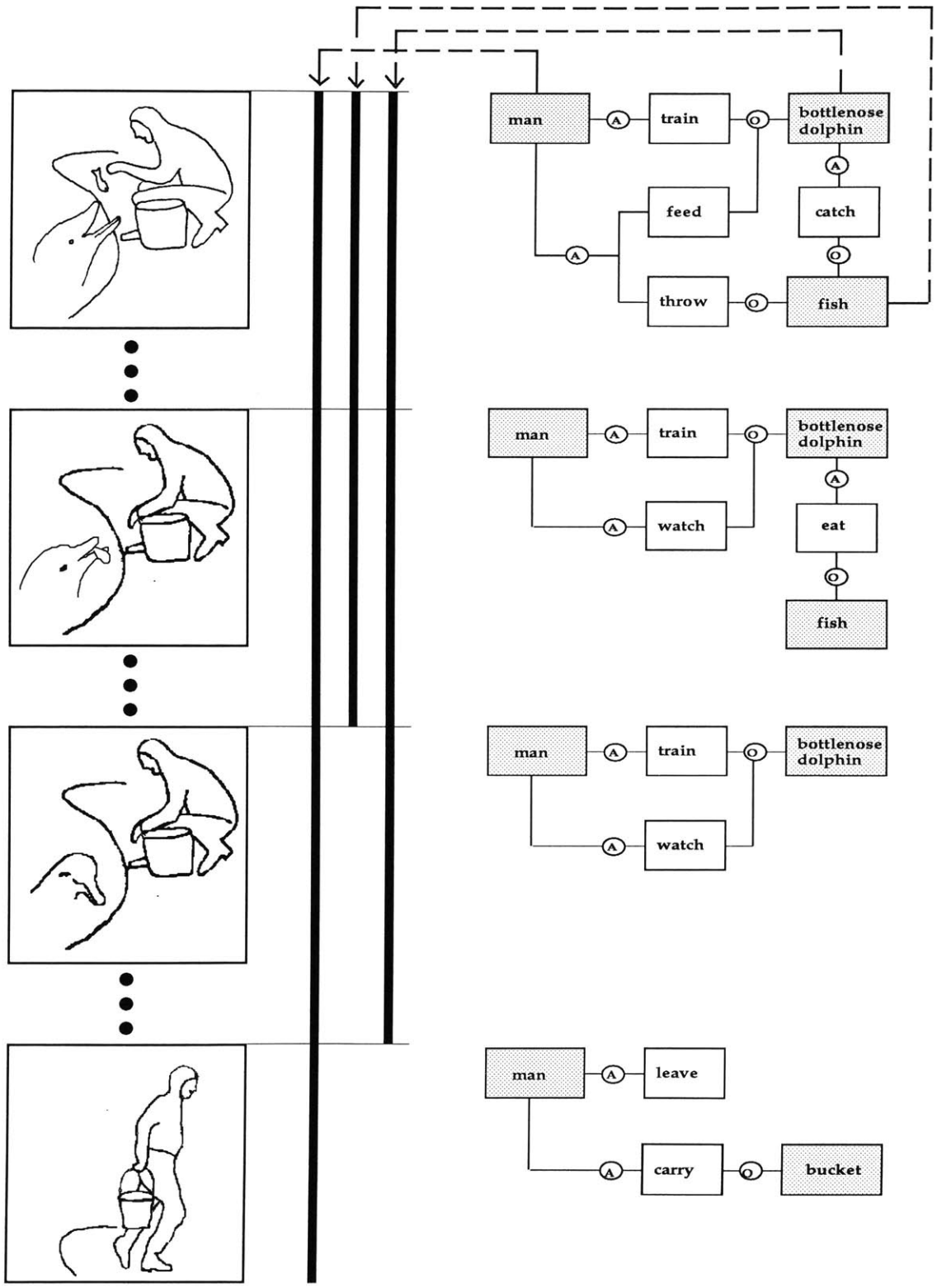


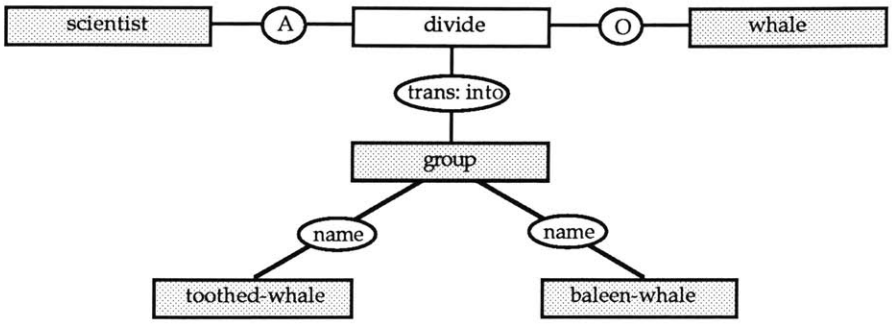
Figure 2.20 Conceptual graphs in a dynamic environment

For example, in figure 2.20, the thick lines indicate the amount of time for which each concept remains visible in the image.

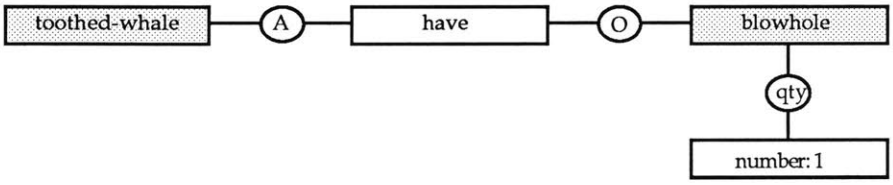
Let us show a few hypothetical situations where the narration track of a video segment can be used to locate related stills or other video segments. The whales videodisc, for example, contains four short motion segments about the different kinds of whales, their characteristics and behavior and their interaction with people. Each segment is separated into short edits corresponding to concise narration statements such as:

1. "Scientists divide whales into 2 groups, toothed whales and baleen whales"
2. "Another feature of toothed whales is that each one has a single blowhole"
3. "All baleen whales have 2 blowholes"
4. "Toothed whales catch fish and squid with their sharp cone-shaped teeth"
5. "Songs of humpback whales are mostly heard in warm waters"

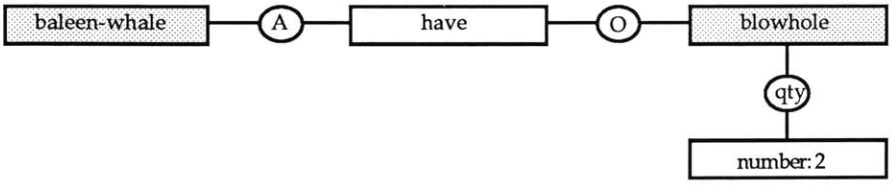
Following this, would be the corresponding graph representations, using Sowa's notation. The graphs represent the narrated statements in the video but may not describe what is actually shown in the image.



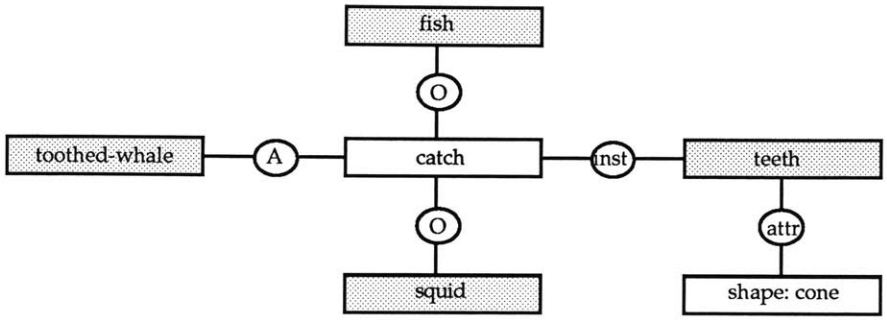
graph #1



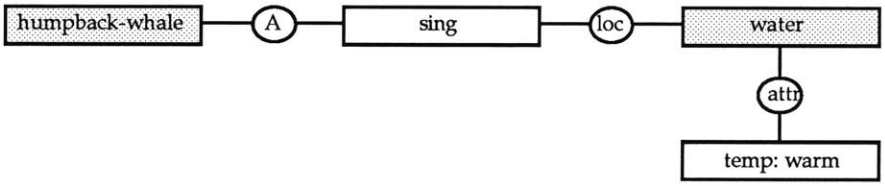
graph #2



graph #3



graph #4



graph #5

Graph #1 can be used to access complementary still images of toothed and baleen whales whereas graph #2 and #3 can find images of blowholes in both whale types. Graph #4 may, for example, point to video segments of dolphins catching fish or perhaps a close-up still of dolphin teeth. Finally, graph #5 would require a sound segment, whether synchronized with the video or by itself.

Sound is very difficult to represent semantically. No system is general enough to cover the various types of sound representations. One idea that would be suitable for the proposed model is to classify sound in terms of *instrument* and *action*. Instruments would correspond to concepts, and actions would correpond to relations. Figure 2.21 shows a simple map between the wave form representation of the sound of a humpback singing and the corresponding conceptual graph.

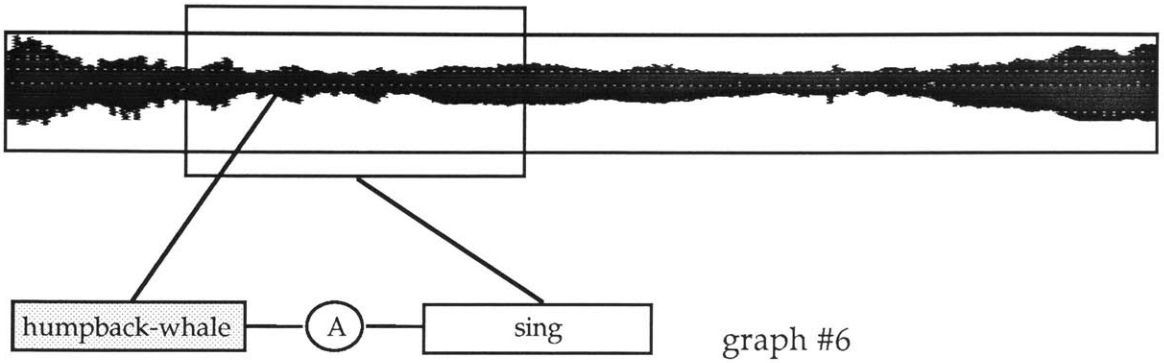


Figure 2.21 A simple graph representation of sound

Graph #6 would then match graph #5.

This representation is over-simplified but is presented here as a pointer for future investigation. Applying this model to dynamic media is a very difficult problem and is currently left open for further research.

CHAPTER 3
IMPLEMENTATION

This chapter describes the implementation of MIDS. The program is written in Lucid Common Lisp using Flavors, an object-based data representation. The object approach allowed for information hiding and abstraction. MIDS uses two main classes of objects: the *image object* and the *concept object.* The first two sections introduce these two classes of objects. Section 3.3 describes the internal construction of the graph referring to the example presented in chapter 2. Section 3.4 describes the association mechanism used by the Matcher.

3.1 The Image Object

The image object uses the following data structure.

```
image-object
                  :topic
                  :subtopic
                  :caption
                  :list-of-regions
                  :list-of-relations
                  :num-regions
                  :num-relations
                  :concept-weights
                  :relation-weights
                  :maximum-relevance
```

The *topic* and *subtopic* slots are only used to maintain the hierarchical ordering of the original material. In the overview example of chapter 2, the topic is "whales" and the subtopic is "human contact". *Caption* is just an extra piece of image information for the designer but is not used by the system. Region objects are created for every defined region in the image. The *list-of-regions* slot contains the names of those objects. *List-of-relations* holds the list of all the agent-to-object relations in the image. *Isa* and *part* relations are kept in the concept object definition. *Num-regions* and *num-relations* keep the number of regions and relations defined in the image. The *concept-weights, relation-weights* and *maximum-relevance* slots are described in section 3.4.

3.2 The Concept Object

Concept objects represent the different concepts in the image. The concept object structure serves two main functions:

- Provide fast access to all associated images by keeping a list of images where it appears.
- Being able to access more general concepts by climbing the *isa* hierarchy.

```
concept-object

                :concept
                :concept-relations

                :is-a
                :level
                :part
```

```
:image-list
```

The *concept* slot holds the name of the concept. The *concept-relations* slot holds the names of the relations between concepts and is used by a heuristic function in the association mechanism. The *level* slot indicates the number of links up the *isa* hierarchy. The *part* slot holds the name of the part belonging to this concept. The *image-list* slot holds the list of images where the concept appears. The *is-a* slot holds the name of the first *isa* object. An *isa* object is created everytime we define a new *isa* relation:

isa-object

```
:is-a
:level

:list
```

For example:

baleen-whale-isa

```
:is-a      whale
:level     1

:list      (blue-whale gray-whale humpback-whale...)
```

whale-isa

```
:is-a      nil
:level     2

:list      (baleen-whale toothed-whale)
```

3.3 The Graph Editor

The implementation is geared towards providing an environment in which the designer who may not be a computer programmer, can easily assign knowledge to images. The construction of the conceptual graph and the association mechanism is therefore transparent to the designer.

The Graph Editor is used to create the semantic descriptor for the image. The latter is constructed graphically by drawing out the regions of the image corresponding to concepts as well as the relations between those concepts.

Let us take the image of section 2.2 and show how the system dynamically creates the different nodes of the graph from the graphical interaction.

3.3.1 Defining Concepts

Since there can be more than one region in the image with the same concept name such as *woman* and *blowhole* (figure 3.2), MIDS allows the designer to label multiple regions, with the same concept name, by using the **define group** command. First, we define the regions for the two women:

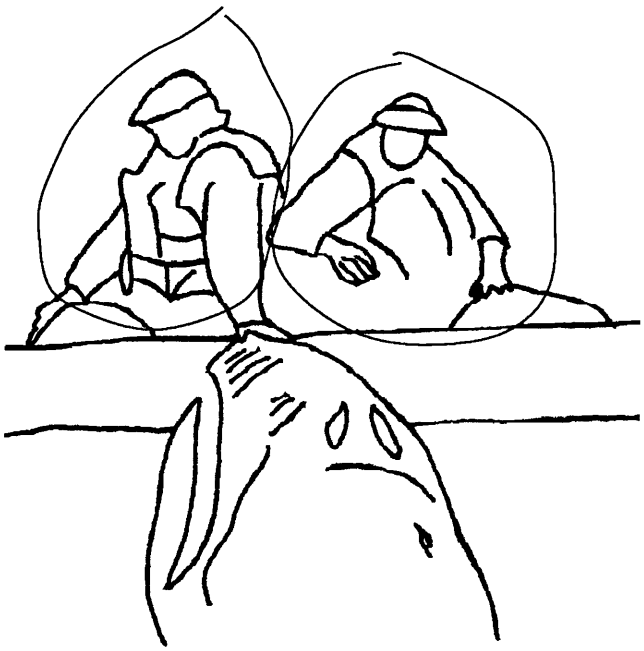


Figure 3.1 Multiple regions can be labelled with the same concept name

The system asks:

name that group

> **woman**

Then, MIDS creates a region object. The name of the region object is created by concatenating the image object name and the region name (in this case: *whaletouch-woman*). *Whaletouch-woman* is added to the slot *:list-of-regions* and *:num-regions* is incremented by 1.

```
whaletouch-image
                    :list-of-regions      (whaletouch-woman*)
                    :list-of-relations    nil
                    :num-regions          1
                    :num-relations        0
```

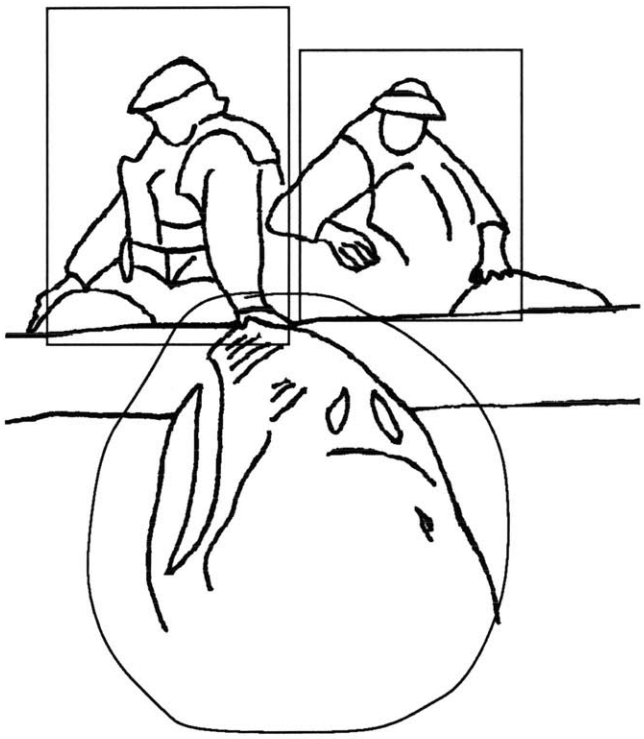
Secondly, the system will check if the concept object for *woman* has been created. If it has, then *whaletouch* gets added to the *:image-list* slot of the *woman* concept otherwise MIDS creates it. In this particular example we will assume that none of the concepts in the image have been created.

```
woman-concept
                    :concept              woman
                    :concept-relations    nil

                    :is-a                 nil
                    :level                0
                    :part                 nil

                    :image-list           (whaletouch)
```

* The bold type indicates new information for the object

The first node of the graph is created:

Next, we define the region for the gray whale.



name that region

> **gray-whale**

The region is added to the list of regions:

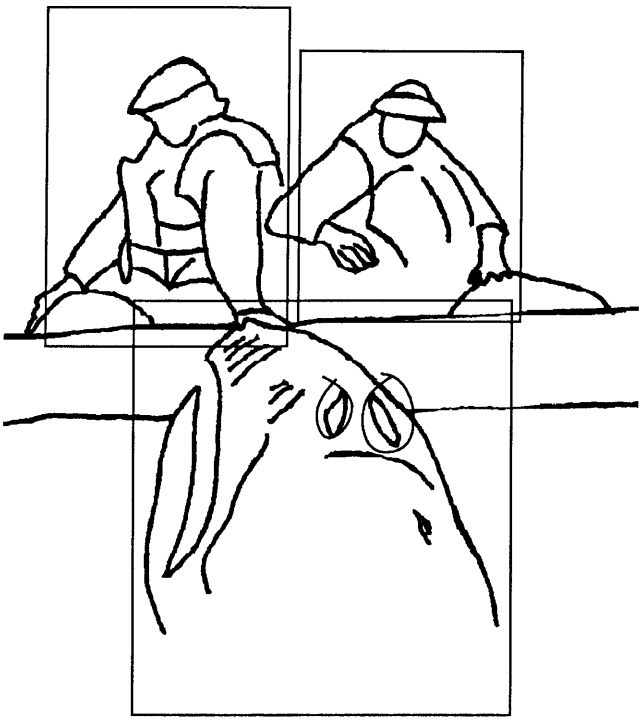
```
whaletouch-image

                    :list-of-regions        (whaletouch-woman
                                             whaletouch-gray-whale)
                    :list-of-relations      nil
                    :num-regions            2
                    :num-relations          0
```

the concept is created:

```
gray-whale-concept

                    :concept                gray-whale
                    :concept-relations      nil

                    :is-a                   nil
                    :level                  0
                    :part                   nil


                    :image-list             (whaletouch)
```

and the graph is updated:

The last regions we will define are the two blowholes of the whale.



name that group

> **blowhole**

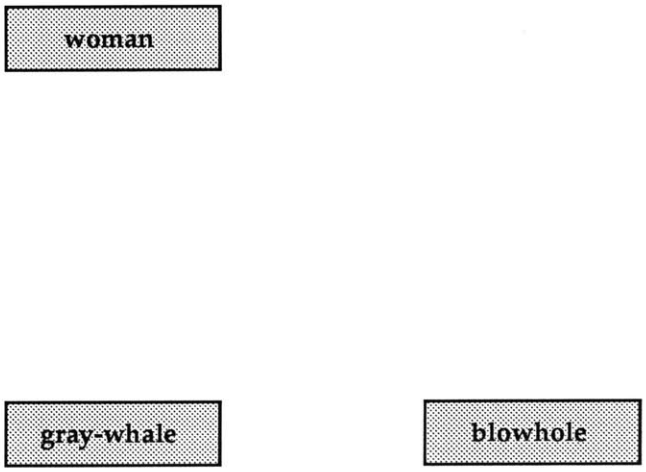
The region is added to the list of regions:

```
whaletouch-image

                    :list-of-regions       (whaletouch-woman
                                             whaletouch-gray-whale
                                            whaletouch-blowhole)
                    :list-of-relations     nil
                    :num-regions           3
                    :num-relations         0
```

the concept is created:

```
blowhole-concept

                    :concept              blowhole
                    :concept-relations    nil

                    :is-a                 nil
                    :level                0
                    :part                 nil


                    :image-list           (whaletouch)
```

and the graph is updated:



3.3.2 Defining Relations

Now that all our regions have been defined, we define the different relations between them. The relations in figure 3.2 are drawn in red. In relation1 the woman is touching the gray whale. In relation2 the other woman is watching the whale. Note that this woman could be directing her right arm towards the whale to touch it too; so relation2 may also be labelled as "touch". However since MIDS does not differentiate between regions with the same

name, labelling relation2 as "touch" would be redundant. Also, introducing a new relation name is always desirable since it augments the association possibilities. Relation3 indicates an *isa* relation. We define the gray whale as being a baleen whale.

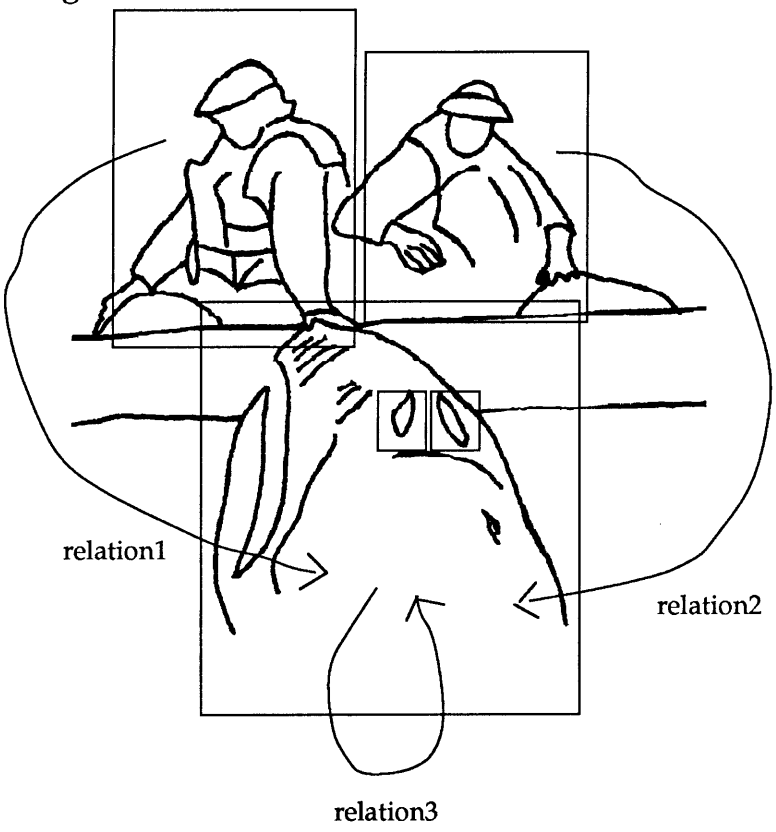


Figure 3.2 Drawing the relations

Relation1 and relation2 are added to the list of relations:

```
whaletouch-image
                    :list-of-regions      (whaletouch-woman
                                           whaletouch-gray-whale
                                           whaletouch-blowhole)

                    :list-of-relations     ((touch whaletouch-woman
                                               whaletouch-gray-whale)
                                            (watch whaletouch-woman
                                               whaletouch-gray-whale))
```

```
:num-regions          3
:num-relations        2
```

Relation1 and relation2 are added to the respective concept:

woman-concept

```
:concept              woman
:concept-relations    (touch watch)

:is-a                 nil
:level                0
:part                 nil

:image-list           (whaletouch)
```

Relation3 is entered in the gray-whale concept object:

gray-whale-concept

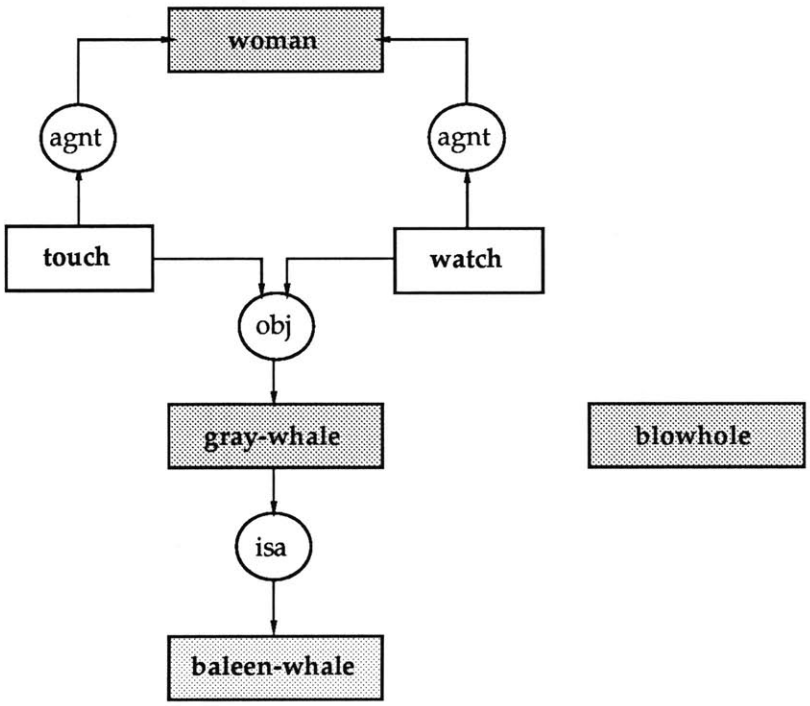
```
:concept              gray-whale
:concept-relations    nil

:is-a                 baleen-whale
:level                0
:part                 nil

:image-list           (whaletouch)
```

The isa-object for baleen-whale is created:

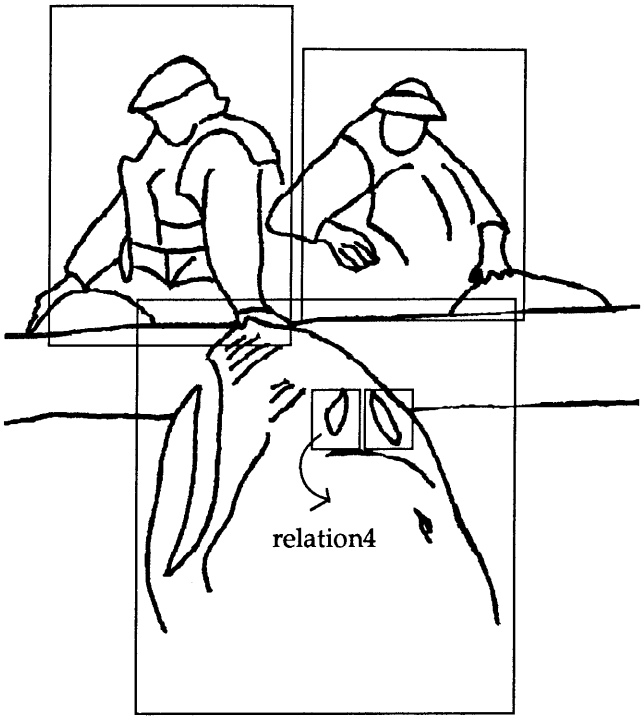
baleen-whale-isa

```
:is-a                 nil
:level                1
:list                 (gray-whale)
```

and the graph is updated:



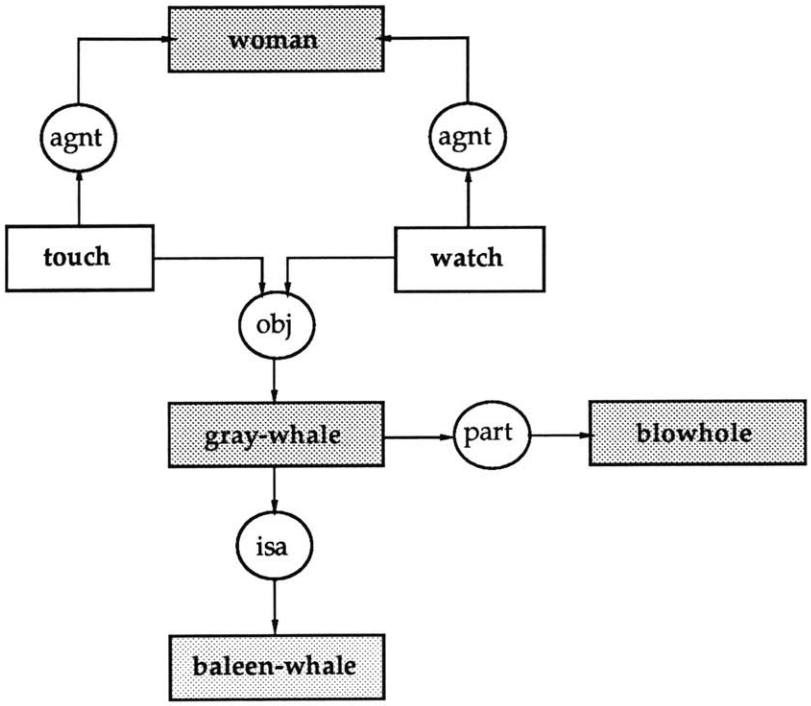
Now we change the color of the sketch to blue to indicate that blowholes are part of the whale.



Relation 4 shows an arrow between the *blowhole* region and the *gray-whale* region defining a *part* relation. The system then automatically stores the relationship into the *gray-whale* object:

```
gray-whale-concept

                    :concept              gray-whale
                    :concept-relations    nil

                    :is-a                 baleen-whale
                    :level                0
                     :part                (blowhole)

                    :image-list           (whaletouch)
```

and the graph is updated:

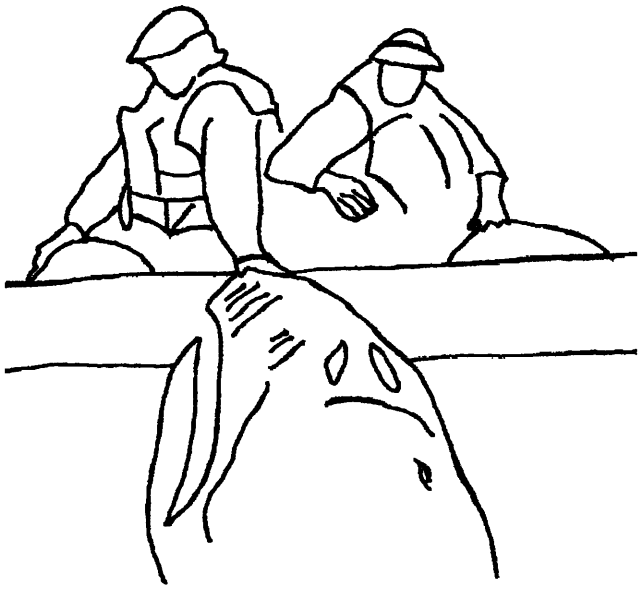


3.4 The Matcher

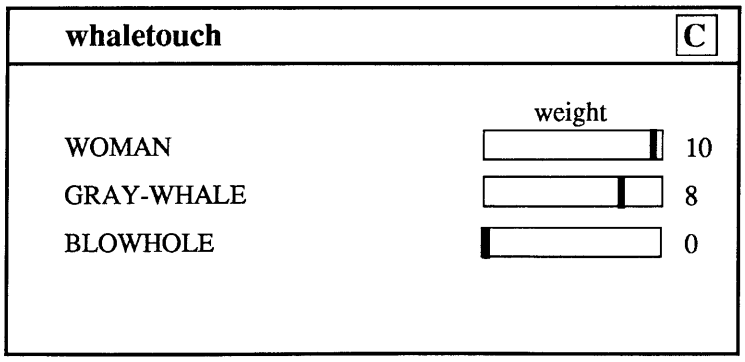
The Matcher compares one image against other knowledgeable images and returns a list of associated images. These images match the original image either through syntactic matching, meaning the graphs share identical nodes or semantic matching by generalizing concept nodes.

3.4.1 The Association Mechanism

We will refer to the scenario presented in chapter 2 to illustrate the association mechanism.



Recall our task of finding other images where people are watching whales. First, we set the weights for the concepts:

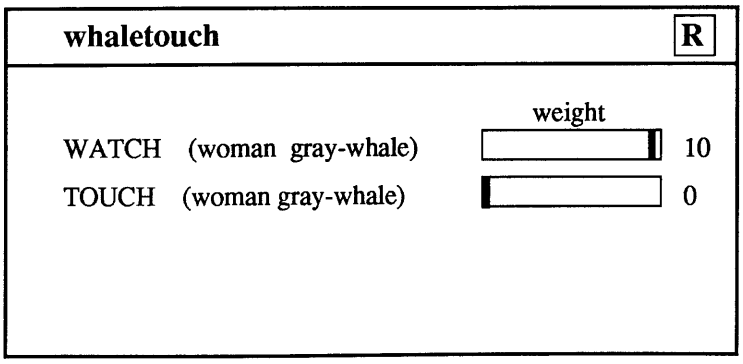


The weight has two interdependent functions. The first one is to calculate the score for each concept based on the user's interest. The second purpose is to determine the order in which the concepts are generalized. The range of the slider is established based on the number of concepts in the image. Since

none of our images contains more than 10 distinct concepts, the range is set to 10 and therefore allows the user to control the generalization order for up to 10 concepts in the image. The weight is then updated in the *concept-weights* slot:

```
whaletouch-image

                    :list-of-regions      (whaletouch-woman
                                           whaletouch-gray-whale
                                           whaletouch-blowhole)

                    :list-of-relations    ((touch whaletouch-woman
                                             whaletouch-gray-whale)
                                           (watch whaletouch-woman
                                             whaletouch-gray-whale))

                    :concept-weights       ((woman 10)
                                           (gray-whale  8)
                                            (blowhole 0))

                    :relation-weights     nil

                    :num-regions          3
                    :num-relations        2
```

we now set the weight for the relation* :



* Unlike for concepts, MIDS does not generalize relations. The relation slider is only used to represent the user's interest.

The weights are then updated in the *relation-weights* slot:

```
whaletouch-image

                    :list-of-regions      (whaletouch-woman
                                           whaletouch-gray-whale
                                           whaletouch-blowhole)

                    :list-of-relations    ((touch whaletouch-woman
                                             whaletouch-gray-whale)
                                           (watch whaletouch-woman
                                             whaletouch-gray-whale))

                    :concept-weights       ((woman 10)
                                           (gray-whale 8)
                                           (blowhole 0))

                    :relation-weights      ((watch 10)
                                           (touch 0))

                    :num-regions          3
                    :num-relations        2
```

All the necessary information has now been recorded into the image object. We can now initiate the search for related images. An assumption is made to give more weight to concept matching than to relation matching. Describing an image in terms of the "objects" it contains generally gives more information than if it was only described in terms of the relations. Hence, a coefficient of 2 is given to matching concepts and 1 to matching relations* .

A maximum relevance value is given to the start image. This value represents the best match of all requested concepts and relations in the image. The maximum relevance for *whaletouch* is:

woman (2*10) + *gray-whale* (2*8) + *watch* (1*10) = 46

* This assumption is discussed at the end of this section

First, all the concepts in all of the images in our scenario, are matched (syntactically) against the ones appearing in the start image. A score is assigned to each concept object based on whether it contains a matching relation or is itself an exact match. This simple heuristic indicates the sequence in which the image lists in each concept object are examined. The total score *S* for a concept object is determined by the following formula:

$$S = m_C * w_C + m_r * w_r$$

where

- m_C represents the matching coefficient for concepts. It is equal to 2 when there is a match and 0 otherwise.
- w_C represents the particular weight for the concept.
- m_r represents the matching coefficient for relations. It is equal to 1 when there is a match and 0 otherwise.
- w_r represents the particular weight for the relation.

Following is the score table for the concepts used in our scenario:

concept	mc	wc	mr	wr	total
woman	2	10	1	10	30
man	0	0	1	10	10
child	0	0	1	10	10
gray-whale	2	8	0	0	16
blue-whale	0	0	0	0	0
humpback-whale	0	0	0	0	0
killer-whale	0	0	0	0	0
bottlenose-dolphin	0	0	0	0	0
risso-dolphin	0	0	0	0	0
fluke	0	0	0	0	0
blowhole	0	0	0	0	0
fin	0	0	0	0	0

The concepts *man* and *child* both get assigned a score since they have a *watch* relation in their *concept-relations* slot:

```
man-concept

                    :concept                man
                    :concept-relations      (watch)

                    :is-a                   person
                    :level                  0
                    :part                   nil

                    :image-list             (museum cornwall
                                            whalewatch seaquarium
                                            killerwhale)

child-concept

                    :concept                child
                    :concept-relations      (watch)

                    :is-a                   person
                    :level                  0
                    :part                   nil

                    :image-list             (childwatch)
```

The Matcher will now look at each concept object by decreasing score and check its *image-list* slot. The graph for each image in the *image-list* slot will be compared with the one from the start image.

A similar scoring system is used to determine the correlation factor between each image in the image list and the start image. The following tables show the matching scores for concepts and relations. We start with the highest-scored concept: *woman*.

concept matching

image	concepts	mc	wc	total
whalewatch	woman	2	10	20
	man	0	0	0
	humpback-whale	0	0	0
				20
cornwall	woman	2	10	20
	man	0	0	0
	risso-dolphin	0	0	0
				20
whaletouch	woman	2	10	20
	gray-whale	2	8	16
	blowhole	2	0	0
				36
flukewatch	woman	2	10	20
	humpback-whale	0	0	0
	fluke	0	0	0
				20

relation matching

image	relation	agent	object	mr	wr	total
whalewatch	watch	woman	humpback-whale	0	10	**0**
cornwall	watch	woman	risso-dolphin	0	10	**0**
whaletouch	watch	woman	gray-whale	1	10	**10**
	touch	woman	gray-whale	1	0	**0**
flukewatch	watch	woman	humpback-whale	0	10	**0**

We then add the 2 totals to obtain the final score for each image. The same score calculation is applied to the other concept objects. The following table summarizes the total score for each image:

```
whalewatch          = 20
cornwall            = 20
whaletouch          = 46
flukewatch          = 20
seaquarium          = 20
childwatch          = 0
killerwhale         = 0
museum              = 0
```

Notice that the highest score is naturally obtained by the start image. The Matcher then compares each score with a relevance threshold value *R*. This value is assumed to be 70% of the maximum relevance of the start image* . Any image whose score is greater or equal to *R* is placed into the result queue. The maximum relevance for *whaletouch* is 46, making *R* approximately 32. This means that no images other than the start image will be returned by the Matcher. Since in our scenario, 7 images were requested, the Matcher will now perform semantic matches on the concepts by generalizing.

Because *gray-whale* has a lower weight than *woman*, it is generalized first. The *isa* slot for *gray-whale* indicates it is a *baleen-whale*:

```
gray-whale-concept

                    :concept              gray-whale
                    :concept-relations    nil

                    :is-a                  baleen-whale
                    :level                0
                    :part                 nil

                    :image-list           (whaletouch)
```

The baleen-whale isa object contains the list of other baleen-whale concepts:

* This assumption is also discussed at the end of this section

baleen-whale-isa

```
:is-a       whale
:level      1

:list       (blue-whale gray-whale humpback-whale...)
```

The Matcher can then substitute any concept in the list for *gray-whale*. However, the weight given to these new concepts is decreased to reflect a lower level of accuracy. The subtracted value is equal to twice the *isa* level. The score table for *woman* is now changed:

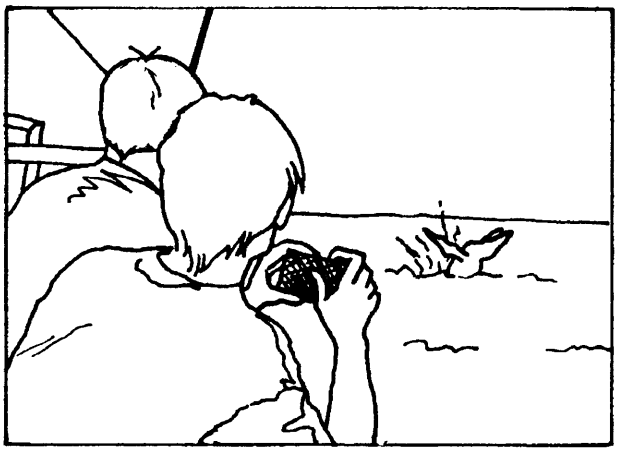
concept matching

image	concepts	mc	wc	total
whalewatch	woman	2	10	20
	man	0	0	0
	humpback-whale	2	6	12
				32
flukewatch	woman	2	10	20
	humpback-whale	2	6	12
	fluke	0	0	0
				32

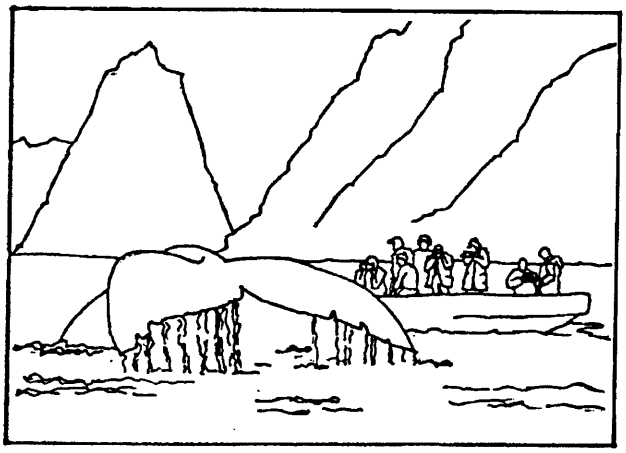
relation matching

image	relation	agent	object	mr	wr	total
whalewatch	watch	woman	humpback-whale	1	8	**8**
flukewatch	watch	woman	humpback-whale	1	8	**8**

The total score for both images is 40, which is greater than R (32). The Matcher will then add *whalewatch* and *flukewatch* to the result queue as the first two images:

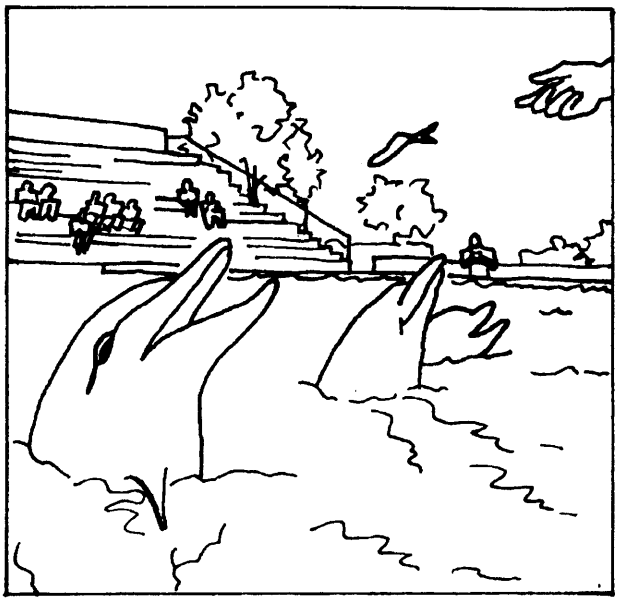


whalewatch

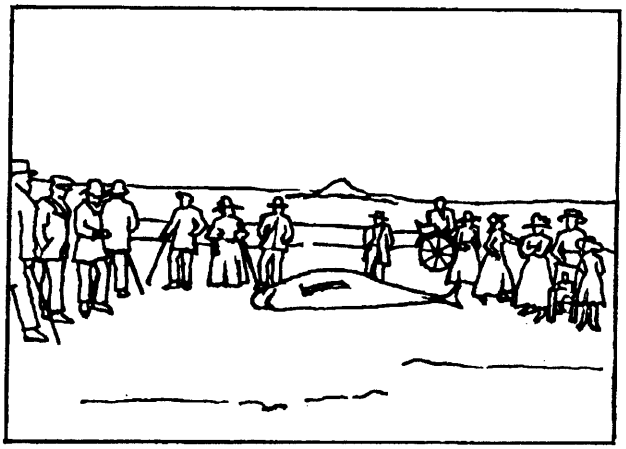


flukewatch

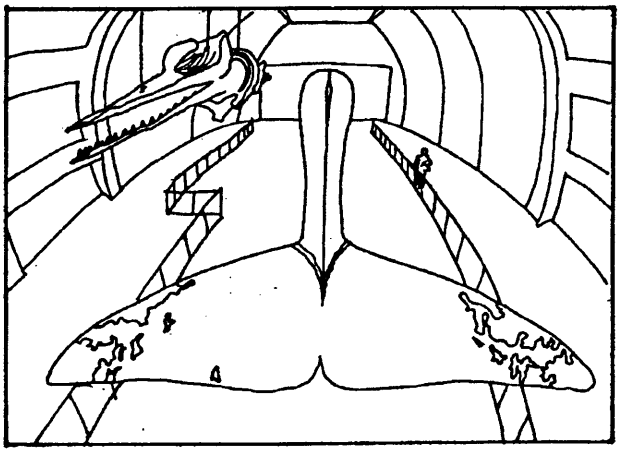
The Matcher will find the other images below in the same fashion:



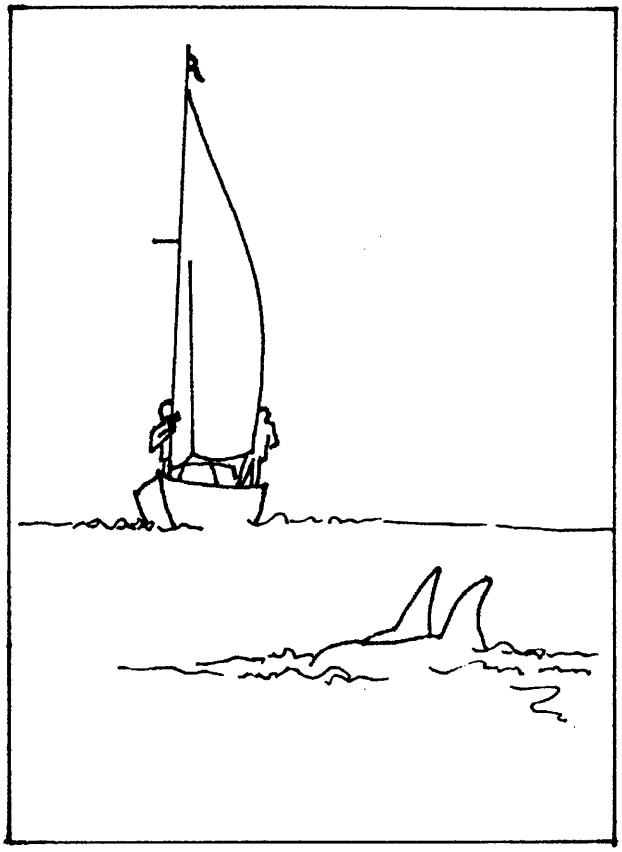
seaquarium



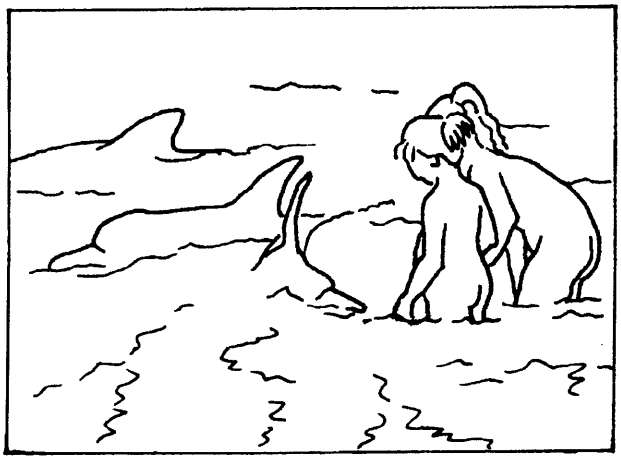
cornwall



museum



killerwhale



childwatch

The implementation of MIDS raises a few questions. First of all, the relevance correlation between two images needs to be investigated. Although some cognitive experiments have examined how people describe single images [Pettersson 88], no examples of the comparison between images have been found.

Secondly, MIDS sets the range of the concept slider to be proportional to the maximum number of concepts that can be defined in an image. This allows the designer to specify a different value for each concept, hence determining the generalization order. Since this value is also the weight or interest that the designer assigns to a particular concept, the question is whether both functions are interdependent or whether a separate slider should be created to control the order of generalization.

Finally, the decrease in weight due to concept generalization should be calculated as part of the concept match coefficient as opposed to being subtracted from the concept weight. The amount by which the concept match coefficient should be decreased also needs to be investigated, although making it proportional to the generalization level seems reasonable.

CHAPTER 4
CONCLUSION

This thesis proposed a model for describing content knowledge in images. The model used conceptual graphs as a way of representing the semantic content of each image object. This representation was then used to associate related images as a result of matching content relations. An interactive system was implemented to allow the information designer to describe the content of images graphically using a free-hand sketch. Once the images are knowledgeable about their content, a multimedia designer can then browse through the images in order to gather material for his/her design.

4.1 Problems with the model

One of the main limitations with this model is the prohibitive expense of manually indexing large numbers of images. Images, unlike text, cannot be automatically indexed. A single subject matter like whales, for example, could easily contain thousands of stills. A solution may be to use a controlled vocabulary for concepts and relations, allowing more than one person to encode content information. However, because the visual information in the image does not change over time, the content knowledge

need not be modified. Another issue to consider is how many concepts and relations need to be represented for an image to be accurately described. Since the image base is created incrementally, it becomes difficult for the information designer to test whether each image has sufficient knowledge about itself to accurately relate to other similar images.

4.2 Future Work

MIDS is only using one type of grammatical structure to describe an action between two concepts, namely a noun phrase construct with a transitive verb. Intransitive verbs are also required to describe images (ie: a *humpback whale* is *diving*). Therefore, the first suggestion for future research is to establish a formal linguistic structure for the categorization of relation types. If other media such as video and sound segments could also be represented in the same fashion, it would be useful to know whether the verbs imply motion (ie: a *bottlenose dolphin* is *jumping*) or sound (ie: a *humpback whale* is *singing*).

Another research direction lies in applying the model to other media. First, the granularity of text, video and sound chunks needs to be established. Whereas an image is a self-contained spatial information unit, text, video and sound segments do not always have clear cut semantic boundaries. The information is spread out in time. It then becomes necessary to assign special

content descriptions or profiles [Salton 89] representing the content of these media segments. These profiles would correspond to individual conceptual graphs and would serve as segment-content surrogates during information retrieval.

Lastly, MIDS could also be enhanced by using a knowledge-base of real world facts. When MIDS does not find any related images it generalizes concepts to provide the user with an alternate set of images. Instead of limiting the failure recovery to concept generalization, the system could use a set of rules about real world facts to explore similar relationships in other images. For example, CYC [Lenat 85], a system being developed at MCC[1], attempts to represent encyclopedic facts such as: "To find food and to navigate through the water, toothed whales use a system called: echolocation." The knowledge representation associated with this fact could be used to find images related to the echolocation system if specific images of dolphins finding food could not be found.

MIDS has only been tested with a small number of relatively simple images. Being able to test the system with a large set of more complex images would better indicate its potential as well as its limitations.

---

[1] Microelectronics and Computer Technology Corporation

BIBLIOGRAPHY

[Akscyn 87] Robert Akscyn, D. L. Mc Cracken, E. Yoder. "KMS: A distributed Hypertext System for Sharing Knowledge in Organizations." Hypertext '87 Papers, November 1987.

[Beauchamp 87] D. Beauchamp. "A Database Representation of Motion Picture Material." Bachelor Thesis, MIT 1989.

[Bloch 87] Gilles, Bloch. "From Concepts to Film Sequences." Short paper, Yale University, Artificial Intelligence Lab, 1987.

[Christodoulakis 86a] Stavros Christodoulakis, F. Ho, M. Theodoridou. "The Multimedia Object Presentation Manager in MINOS: A Symmetric Approach." Proceedings, ACM SIGMOD, May 1986.

[Christodoulakis 86b] Stavros Christodoulakis, Christos Faloutsos. "Design and Performance Considerations for an Optical Disk-Based, Multimedia Object Server." COMPUTER, December 1986.

[Conklin 87] Jeff Conklin. "Hypertext: An introduction and Survey." COMPUTER, September 1987.

[Dondis 73] Donis A. Dondis. "A primer of Visual Literacy." MIT Press, 1973.

[Flight 89] John L. Flight. "A System for Recognizing Hand Drawn Marks." Bachelor Thesis, MIT, 1989.

[Greenlee 88] Russell L. Greenlee. "From Sketch to Layout: Using Abstract Descriptions and Visual Properties to Generate Page Layout." Masters Thesis, MIT, 1988.

[Hodges 89] Matthew E. Hodges, Ben H. Davis, Russell M. Sasnett "Investigations in Multimedia Design Documentation." in "The Society of Text" MIT Press, 1989.

[Hooper 88] Kristina Hooper. "Interactive Multimedia Design 1988." The Multimedia Lab, Apple Computer, Inc. Technical Report #13, November 1988.

[Lenat 85] Douglas B. Lenat, Prakash, Mayank, Shepherd, Mary. "CYC: Using Common Sense Knowledge To Overcome Brittleness and Knowledge Acquisition Bottlenecks." MCC Technical Report AI-055-85 - July 15, 1985.

[Levitt 88] David Levitt, Glorianna Davenport. "Symbolic Description of Movie Media." Media Laboratory, MIT, Proposal to the National Science Foundation, 1988.

[Meyrowitz 86] Norman Meyrowitz. "Intermedia: The Architecture and Construction of an Object-Oriented Hypertext/Hypermedia System and Applications Framework." Proceedings, OOPSLA, September 1986.

[Pettersson 88] Rune Pettersson. "Visuals for information" Esselte Forlag. 1988.

[Quillian 68] Ross M. Quillian. "Semantic Memory" in "Semantic Information Processing", MIT Press, 1968.

[Salton 89] Gerard Salton "Automatic Text Processing" Addison-Wesley, 1989.

[Sasnett 86] Russell M. Sasnett. "Reconfigurable Video" M.S.V.S Thesis in Architecture, MIT, 1986.

[Shank 72] Roger C. Shank. "Conceptual dependency: a theory of natural language understanding." Cognitive Psychology, *3* 552-631

[Shastri 88] Lokendra Shastri. "Semantic Networks: An Evidential Formalization and its Connectionist Realization."

[Shlaer 88] Sally Shlaer, Stephen J. Mellor. "Object-Oriented Systems Analysis" Yourdon Press, 1988.

[Sowa 84] John F. Sowa. "Conceptual Structures - information processing in mind and machine." Addison-Wesley, 1984.

[Wilson 87] S. K. Wilson. "Palenque: An Interactive Multimedia Optical Disk Prototype for Children." Bank Street College of Education, New York, Technical Report #2, 1987.

[Woods 75] William A. Woods. "What's in a Link: Foundations for Semantic Networks." Reading in Knowledge Representation, editors Ronald Brachman, Hector Levesque. Morgan Kaufmann, 1975.

Image Sources

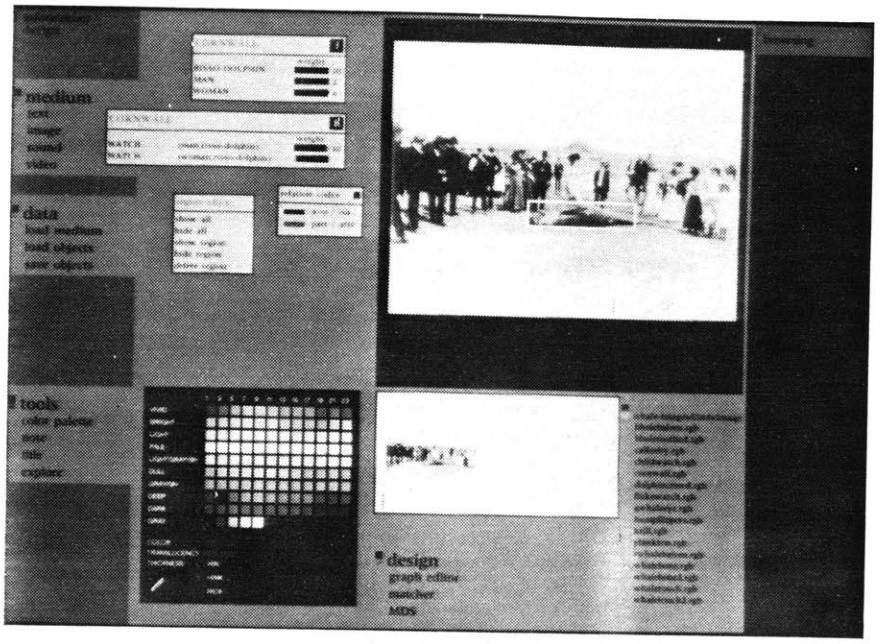
[National 81] "WHALES" - Nebraska Videodisc Design/Production Group and National Geographic Society.

[Sierra 83] "The Sierra Club Handbook of Whales and Dolphins." Sierra Club Books, 1983.

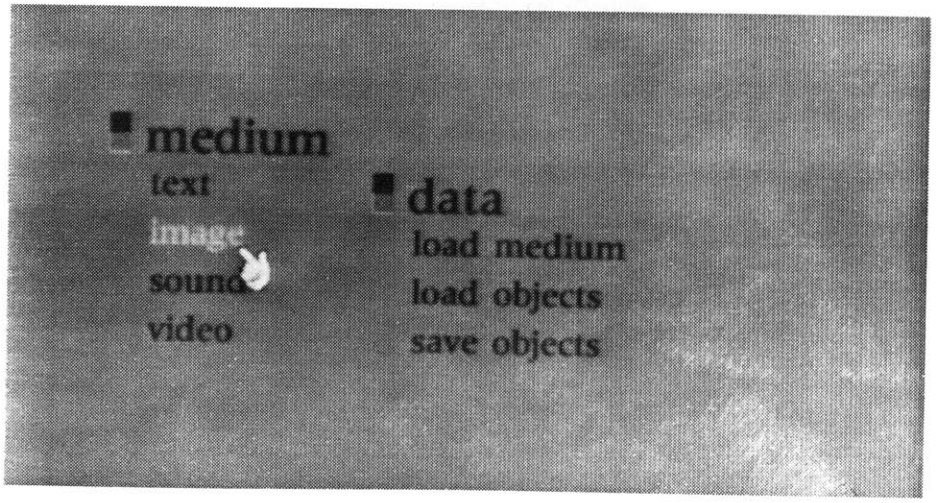
[Intercontinental 88] "Whales, Dolphins and Porpoises" Intercontinental Publishing Corporation, 1988.

APPENDIX A

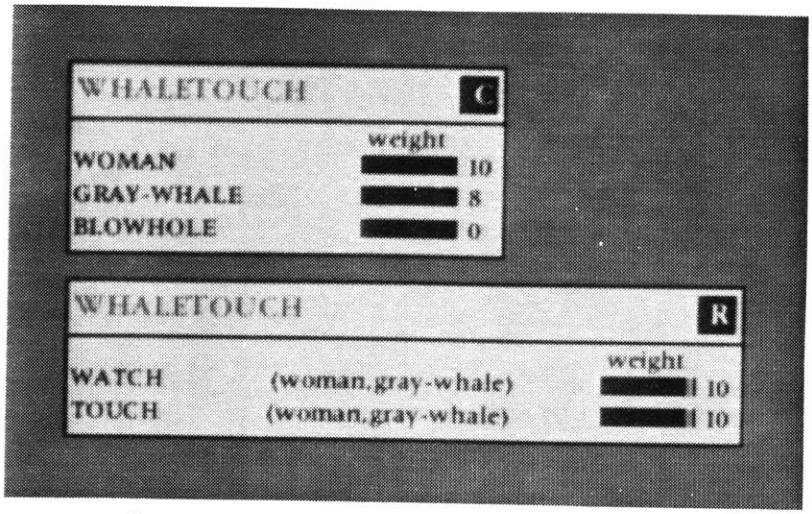
The Multimedia Information Design System interface



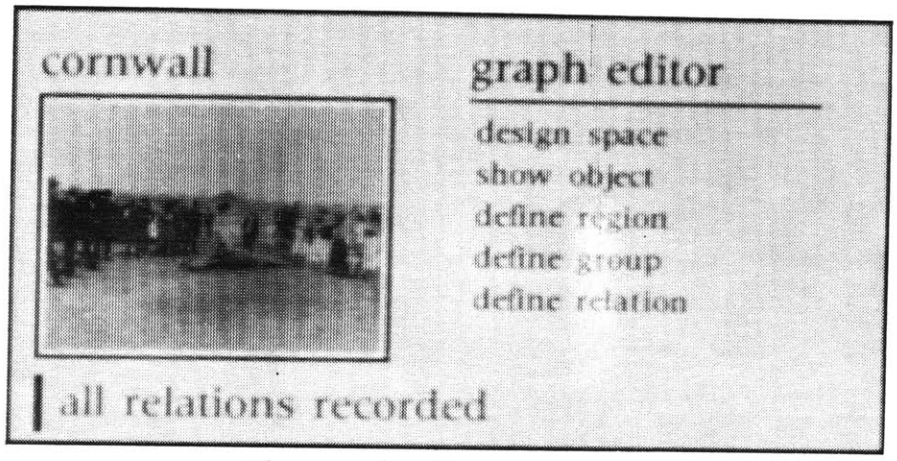
The Interface



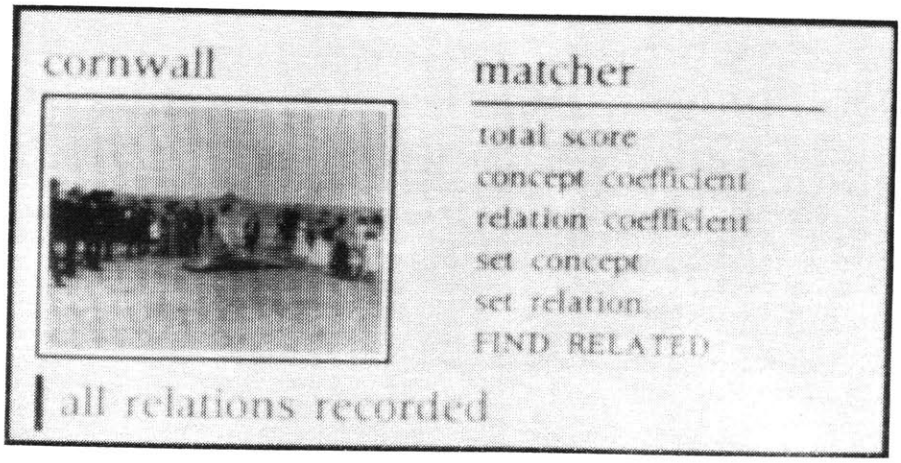
The data menu



Concept and relation interest menus



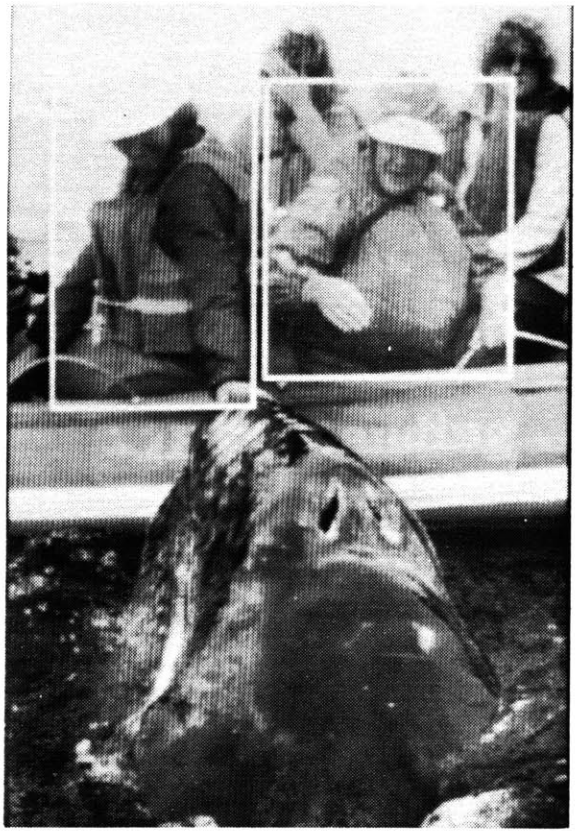
The graph editor menu

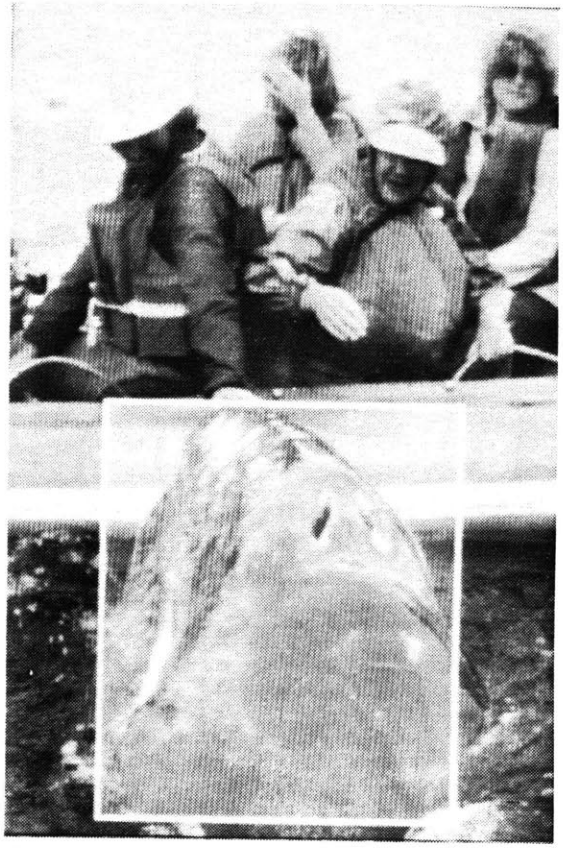


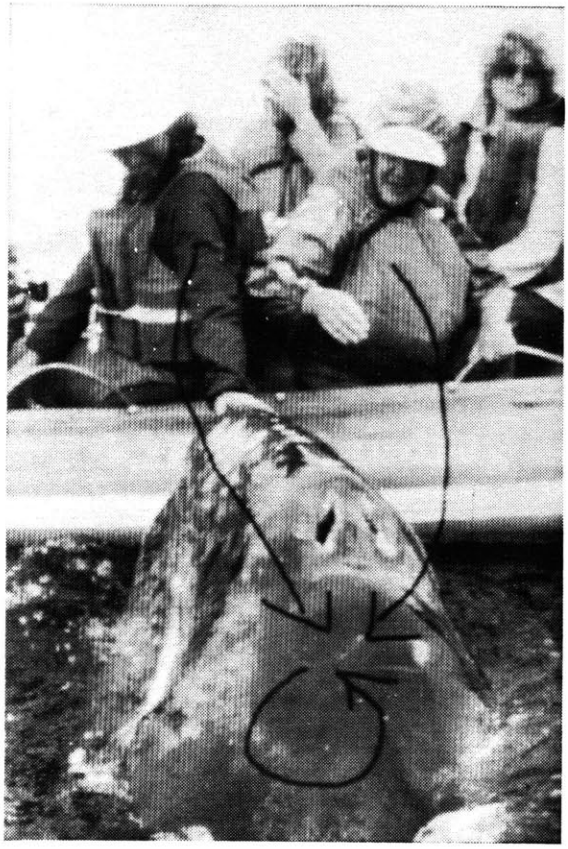
The matcher menu

APPENDIX B

Images with regions defined







Defining relations

APPENDIX C

Glossary

conceptual graph
a graph representation of sentential forms. Conceptual graphs are generally used to assert a single proposition.
(i.e. "The dolphin eats the red fish.")

concept object
element of a multimedia object.
(i.e. "dolphin" is a *concept object*)

dynamic medium
a medium whose display form changes over time.
(i.e. video, sound)

electronic multimedia
using multiple media in an electronic environment.

hypermedia
accessing:
- more information using the same medium (i.e. hypertext.)
- more information using a different medium.

knowledgeable media
multimedia objects which have been assigned a semantic descriptor.

multimedia document
interactive or passive display of multimedia information.
(i.e. an electronic book)

multimedia information
information that can be conveyed with more than one medium.

multimedia information design
the process of creating knowledgeable media and the evaluation of the associative retrieval of those objects as a result of matching their semantic descriptors.

multimedia object
element of a multimedia document.
(i.e. a still is an *image object*)

semantic descriptor
conceptual graph(s) representing the semantic content of a multimedia object.

static medium
a medium whose display form does not change over time.
(i.e. text, still images)

"I miss the good old days when all we had to worry about was nouns and verbs."