

# An evaluation of semantic fisheye views for opportunistic search in an annotated image collection

Paul Janecek, Pearl Pu

Human Computer Interaction Group, I&C-IFF, Swiss Federal Institute of Technology (EPFL), 1015 Lausanne, Switzerland  
e-mail: {paul.janecek,pearl.pu}@epfl.ch

Published online: 15 February 2005 – © Springer-Verlag 2005

**Abstract.** Visual interfaces are potentially powerful tools for users to explore a representation of a collection and opportunistically discover information that will guide them toward relevant documents. *Semantic fisheye views* (SFEVs) are *focus + context* visualization techniques that manage visual complexity by selectively emphasizing and increasing the detail of information related to the user's focus and deemphasizing or filtering less important information.

In this paper we describe a prototype for visualizing an annotated image collection and an experiment to compare the effectiveness of two distinctly different SFEVs for a complex opportunistic search task. The first SFEV calculates relevance based on keyword-content similarity and the second based on conceptual relationships between images derived using WordNet. The results of the experiment suggest that semantic-guided search is significantly more effective than similarity-guided search for discovering and using domain knowledge in a collection.

**Keywords:** Context+focus techniques – Semantic fisheye views – Visual information retrieval interfaces – Empirical user study

## 1 Introduction

Search is often iterative, interactive, and opportunistic; analyzing the results of a query often leads to the discovery of unfamiliar vocabulary and relationships that guide the future direction of search. Researchers have identified a wide range of organizational structures [11, 25] and strategies [1] useful for effective information seeking. In this paper we present an experiment to evaluate an interactive visualization technique that visually integrates these structures and strategies into the interface to guide a user over a collection of information.

Visual information retrieval interfaces (VIRI) use information visualization techniques to reveal relationships between documents and allow a user to rapidly shift between search and browsing tasks [9]. The tight coupling between visualization and interaction make VIRIs potentially powerful tools for discovering global relationships between documents. However, there are several properties of visual representations that limit their effectiveness for opportunistic search. First, as the amount and complexity of information accessible in a VIRI grows, it is increasingly difficult to represent all of it in the limited space of a display. A visualization is often a compromise between showing a small amount of information in detail or a large amount of information abstractly. Second, within a dense display of information, it is often difficult for users to see or visually distinguish the subset that is immediately useful for their current task. Selecting or navigating to each item to view detailed information incurs a cost in time and cognitive effort. Third, any single visual representation optimally supports only a small set of tasks [5, 28]. Supporting a diverse set of tasks, such as comparing content and structural and temporal relationships between documents, will often require multiple representations.

*Semantic fisheye views* (SFEVs) are interactive *focus + context* visualization techniques that are designed to address these limitations by monitoring and revealing the relative importance of information for the user's current task [8]. SFEVs emphasize or increase the detail of the most important information and deemphasize or filter less important information [13]. This is useful for both selecting the information to represent in the constraints of a single display and reducing the visual complexity within a dense visualization. Furthermore, by monitoring relative importance in multiple contexts, these techniques can interactively reveal semantic or structural relationships that may not be visible in the original representation.

We have developed a prototype that uses SFEV techniques to support a number of different opportunistic search strategies within a large collection of professionally annotated images. The first technique emphasizes images and keywords that are similar in content to the focus, and the second technique emphasizes information that is conceptually related to the focus based on WordNet, a general lexical ontology of the English language [17]. These different SFEVs correspond to classic search strategies used for opportunistic search over heterogeneous collections of information [1, 2]. The prototype demonstrates how SFEVs can be used to support a variety of different search strategies.

We have conducted an experimental evaluation comparing the effectiveness of similarity and semantic metrics for complex, iterative search tasks. The initial results of the experiment suggest that semantic-guided search is more effective than similarity-guided search for these tasks.

In the following section we briefly introduce *focus + context* visualization techniques. We then describe our SFEV prototype, which uses similarity and semantic relationships to guide search over an annotated image collection. The framework this prototype is based on is independent of any particular representation and coordinates the multiple interest metrics and emphasis techniques that create the fisheye view effect. We then discuss the design and results of the experimental evaluation.

## 2 Related work

### 2.1 Focus + context techniques

Furnas first described fisheye views as a technique for selectively reducing the information in a display to show the most interesting items, where the *degree of interest* (DOI) was calculated as a tradeoff between a priori importance (global context) and relevance to the user's current task [8]. Furnas suggested that this general technique could be used to create compact views in a variety of different domains by redefining the function that calculates DOI.

Since then, researchers have developed a wide range of fisheye view or *focus + context* interactive visualization techniques. Distortion techniques [15] use geometric transforms to magnify the area spatially near the focus in the view. Graphical fisheye views [26] achieve similar effects by increasing the size or detail of information related to the focus within the structure and constraints of a graphical model. The effectiveness of both distortion techniques and graphical fisheye views for complex search tasks depends largely on whether distance within the view or graphical model corresponds to the needs of the user.

*Semantic fisheye views*, on the other hand, are independent of a particular visual representation and cal-

culate *conceptual distance* from the focus within one or more related data models [13]. Several other researchers are investigating *focus + context* techniques that are based on structural or semantic relationships rather than a particular view. For example, ScentTrails [19] calculates the DOI of objects in a Web page using a model of Information Scent [6, 21] and then highlights relevant information by modifying the underlying HTML.

### 2.2 Exploring image collections

We refer to opportunistic search as an iterative, interactive process that evolves in response to the information found and that encompasses both directed search and browsing [2]. Browsing behaviors with different goals, strategies, and at different levels of granularity are a part of this process [16]. Furthermore, the results of information seeking are not limited to documents but also include the *knowledge* accumulated during the search process [2]. This knowledge is essential for understanding and using the information discovered during search [25].

Several researchers have investigated visualizing image collections in maps based on visual features such as color, texture, and shape [20, 24]. Our prototype uses a similar layout but positions images based on text annotations. In a study of the effectiveness of these layouts for search, Rodden et al. found that users preferred organizing and searching for images based on keywords [24]. PhotoMesa [3] organizes collections of images hierarchically by their metadata and visualizes them in dense space-filling maps. Users can rapidly browse over the images using sophisticated zooming techniques. A significant difference between this and the SFEV approach is that the zoom effect is not applied uniformly over the entire collection but rather to each image separately based on a measure of its relevance. This allows a user to rapidly see and access images that are not spatially near each other. Another very significant difference is that our approach integrates a semantic model directly into browsing.

Other researchers have also developed prototypes that allow semantic search and browsing. Hollink et al. [10] recently developed a prototype that integrates four ontologies for searching within and annotating an image collection. Yee et al. [27] developed and evaluated a browser that organized an image collection by concept hierarchies and allowed users to interactively navigate over the network of relationships between them. Both of these prototypes are hypertext browsers. Our prototype, on the other hand, uses visualization and interaction techniques to allow users to explore semantic relationships. In the comparative study of browsing techniques conducted by Yee et al., they found that users strongly preferred this semantic search to a more classic keyword-based search. The experiment that we describe in this paper also compares keyword-based and semantic browsing, but within the same type of interactive visual environment.

### 3 A semantic fisheye view prototype

To evaluate the effectiveness of SFEVs for opportunistic search, we have developed a prototype that allows a user to interactively explore a large collection of images and keywords using two different types of search strategies. The first SFEV is based on the similarity and direct relationships between objects in the collection, which corresponds to traditional information retrieval search and browsing behavior. The second SFEV uses WordNet to find images and keywords that are conceptually related to the focus. When the user’s focus changes, the prototype recalculates the DOI of related objects using a combination of different metrics and then updates their representation using visual emphasis techniques.

The prototype allows a user to search for images using keyword-based and concept-based queries and rapidly browse over the results by simply brushing over the visual representation of images and their keywords. As the focus changes, the interface calculates the relative interest of all objects in the workspace and smoothly animates changes in their representation. The cost of accessing information in the local workspace is much lower than query-based interaction, which encourages opportunistic search and exploration. In the following sections we describe the prototype in greater detail.

#### 3.1 Software architecture

The prototype is a Java application that integrates three main software components: two packages that manage all queries, visualization, and interaction with the image and WordNet databases, and *VisAmp*, a general library that manages the interest metrics and emphasis techniques used by SFEVs. The *VisAmp* library was originally developed as part of a prototype for browsing tabular information [13, 23] but is independent of the representation and datasets. We based parts of the architecture on several abstract classes used for general layout and display of graphs in the *graphview* library originally developed for the KAON project [18].

#### 3.2 The Corbis image collection

The prototype allows the user to browse over a diverse collection of over 56 500 images that Corbis Corp. made accessible to us for this research. The images in the collection cover a wide range of subjects and are very well annotated with a rich vocabulary of over 28 000 unique keywords. Each image in the collection was professionally annotated by Corbis with keywords describing various aspects of the image, such as people, animals, objects, actions, mood, and location. Each image has an average of 23 keywords, which is sufficient to enable use of traditional information retrieval techniques with few modifications.

An important property of the keyword annotations is that they were produced with the aid of a propri-

etary thesaurus that is not distributed with the images. This improves the quality of the annotations by ensuring a controlled vocabulary and expands each concept with both a limited number of common synonyms. In addition, the thesaurus allows the annotators to add a list of related keywords that represent the path upward to progressively more general concepts. From a search perspective, this allows a user to find an image when their search was for a more general concept. From a browsing perspective, this allows a user to learn the vocabulary of the classification hierarchy and navigate to more general concepts.

#### 3.3 Visual layout

We use a spring layout to position the images and keywords in the workspace by modeling both types of objects as nodes in a graph and the relationships between them as edges. The importance of a keyword in an image is assigned to the weight of the edge connecting them. A single keyword may be connected to multiple images, and there are no direct connections between images or between keywords. The spring layout tends to place highly connected sets of nodes in clusters, and minimally connected nodes are moved to the outside of the visualization.

The implementation of the spring layout is adapted from the *graphview* package of the KAON library [18]. In general, a spring algorithm assigns a repulsive force to the nodes in a graph and models the edges between nodes as springs with a tension based on their weight. To reduce empty space in the layout, we separated the repulsive force of each node into  $x$  and  $y$  components that varied according to the image or keyword dimensions. Furthermore, to increase the legibility of the keywords, we minimized the vertical overlap by exaggerating their  $y$  component of repulsion. The algorithm initially places nodes randomly and then iteratively attempts to minimize tension. The resulting layout tends to place highly connected sets of nodes into clusters and minimally connected nodes toward the outside of the representation. Figure 1 shows the effect of applying this algorithm to a query for the concept “mares.” Words that are often used in common will tend to be positioned near each other, e.g., “colt” and “foal” in the bottom right of the figure or “maternal” and “motherhood” in the top left.

#### 3.4 Semantic fisheye views

Interaction with the image collection is based on our SFEV framework [13]. We use the following general equation to calculate the DOI:

$$DOI_{\text{context}}(x|fp = y) = API(x) + \sum_{i=1}^n w_i dist_i(x, y) . \quad (1)$$

This equation identifies two components that influence the DOI assigned to each object,  $x$ , in the interface



**Fig. 1.** Images and keywords loaded into the workspace are positioned using a spring layout algorithm. The visual weight of the images reflects their relevance to a query for the “mare, female horse” concept<sup>1</sup>

given the current focus,  $fp = y$ . The first is a priori interest (API), which models the relative importance of objects independently of user focus. Objects with high API are landmarks in the information collection. For example, “mares” and “horses” are shown prominently in Fig. 1. The second component models the importance of objects with respect to the current focus and is the result of one or more interest metrics that measure the distance between each object and the focus in a dimension of interest,  $dist(x, y)$ . The relative contribution of interest metrics can be adjusted with a weight,  $w$ . The equation can be adapted to model different contexts of interest, such as keyword or conceptual similarity, by modifying the components. Each distance metric calculates relative interest in a different way, and potentially within different semantic models. We use different combinations of interest metrics to support the two different browsing techniques in the prototype.

### 3.5 Similarity-guided browsing

The first set of interest metrics are designed to reveal information in the collection that is similar to the current focus based on content. We derive the similarity between keywords and images from the order in which keywords are used to annotate an image. This heuristic obviously depends on how the images are annotated but generally works well in this collection as an indirect result of how Corbis uses a thesaurus to create their annotations.

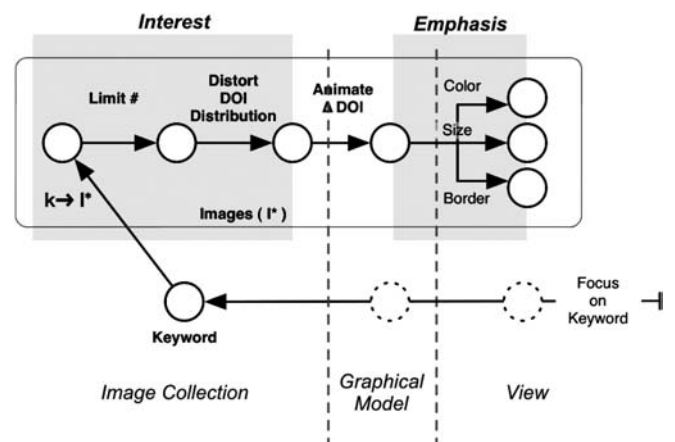
<sup>1</sup> The color version of this paper is available at <http://hci.epfl.ch/publications/2004/janecek-jd1104.pdf>.

Figure 2 shows a model of similarity-guided search when a keyword is the focus. The data state model traces the flow of information from the user’s interaction on the bottom right through the similarity metric and back to the updated view. Each node in the model is a distinct state of the data, and the edges represent transforms between states. The model is divided into three vertical regions representing different stages in creating a visualization: the *Image Collection*, the *Graphical Model*, and the *View*. When the user focuses on the representation of a keyword node in the view, the prototype traces the node back to the underlying keyword object in the image collection and passes this to the similarity metric.

The links from the keyword to related images are then expanded, and this collection is limited to an ordered set of the most important ones, based on the weight of the links between the images and keywords. The distribution of DOI values is distorted and scaled to increase contrast and then passed to a function that animates changes in the DOI of objects in the collection. Emphasis techniques depend on the DOI value to determine the visual weight to render objects, so any number of visual properties can be automatically coordinated in this way. Figure 3 shows the similarity value calculated using this metric for an image to all of its keywords. The degree of interest is shown using size and saturation in the representation of the images, keywords, and edges.

A similar process is followed to calculate the similarity between images, as shown in Fig. 4, with the resulting screenshots shown in Fig. 5. When the limits of these metrics are changed, the prototype can be configured to progressively add information to the display in descending order of DOI.

The metrics discussed above reveal similarities between objects and create visual clusters of tightly connected nodes. When the user changes focus, the interface recalculates the DOI of every object to reflect their similarity to the new focus. This supports opportunistic discovery of images that are similarly annotated and the keywords that are used in the vocabulary of the col-



**Fig. 2.** Calculating the similarity between keywords and images

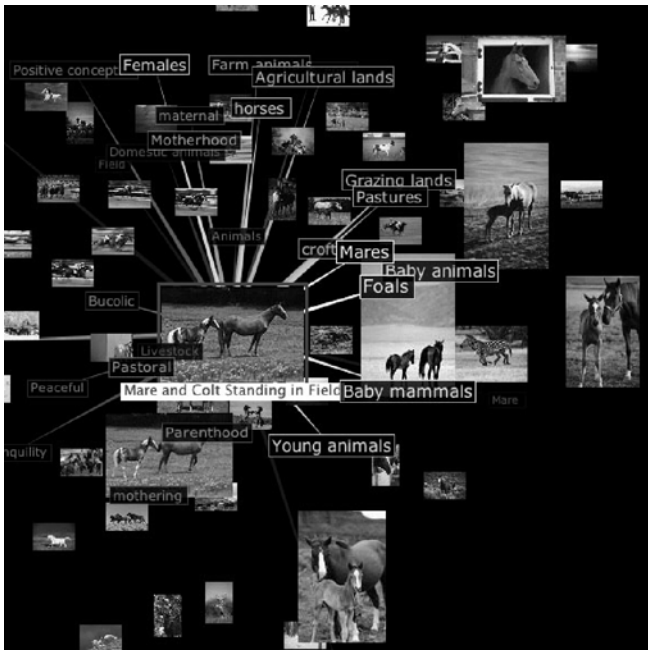


Fig. 3. Representing the similarity between an image and its related keywords

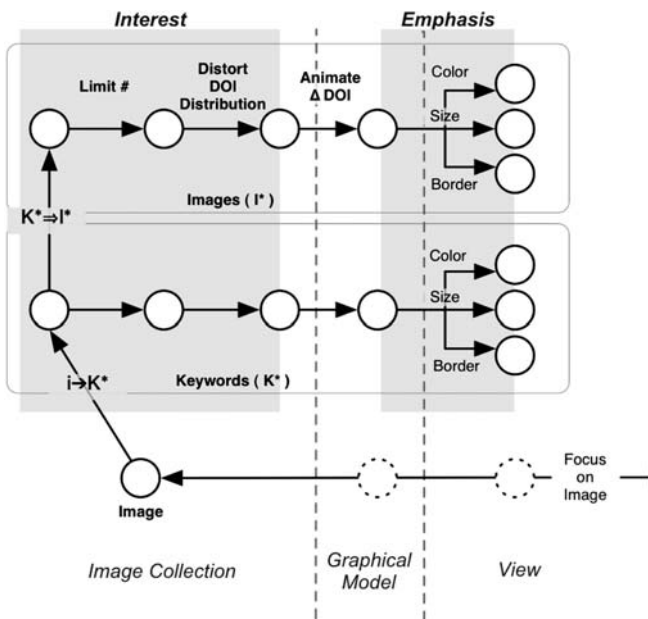


Fig. 4. Calculating the similarity between images

lection. In the following section, we describe metrics for finding information that is conceptually related but not similarly annotated.

### 3.6 Semantic-guided browsing

The second type of SFEV implemented in the prototype allows a user to search and browse over the image collection using the semantic relationships modeled in WordNet. In the next section we give a brief overview of WordNet and then describe our implementation.

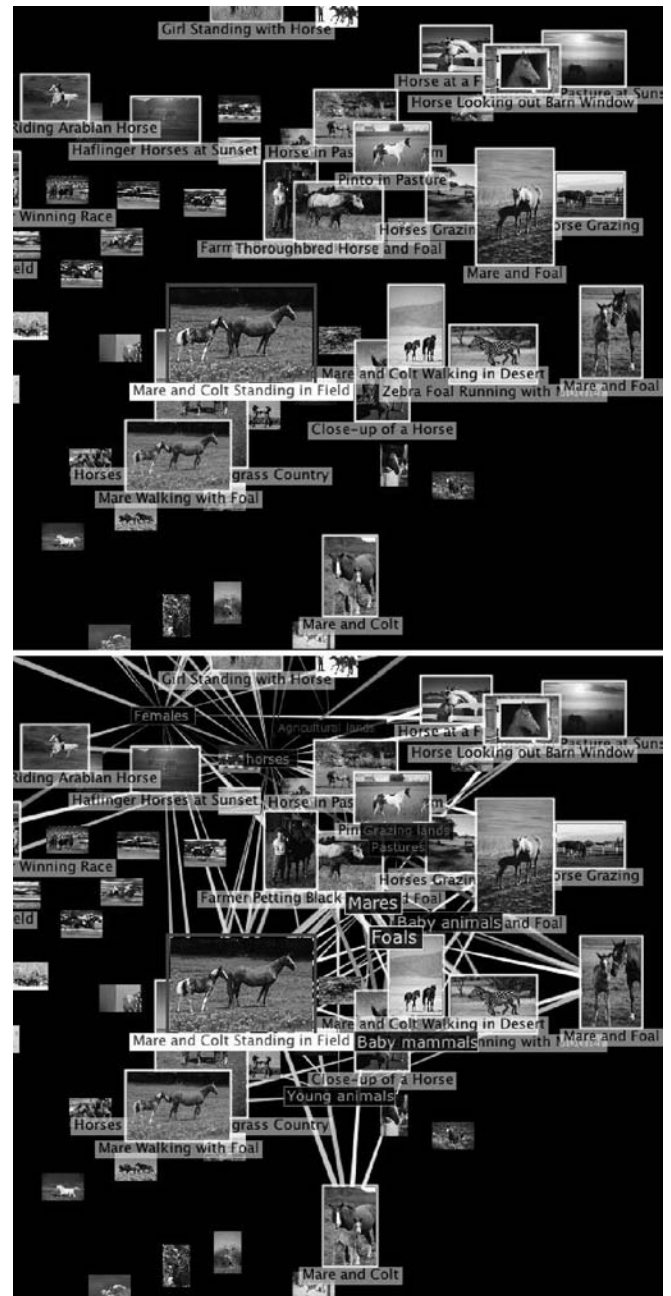


Fig. 5. Representing the similarity between images. The top screenshot uses image size to indicate similarity. The bottom screenshot adds a small number of important keywords relating the images

#### 3.6.1 WordNet

WordNet is a general lexical ontology originally developed at Princeton University in 1985 to test psycholinguistic theories of the English language [17]. WordNet has continued to evolve since then, and the version this research is based on (1.7.1, August 2002) contains approximately 140 000 unique word forms with 111 000 different senses. Only the base forms of words are stored in the WordNet database and are referred to as *lemmas*. The different senses of words (i.e., concepts) are modeled as

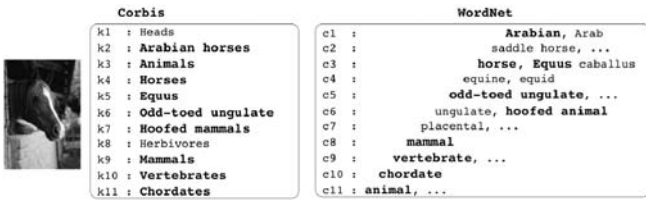


Fig. 6. Correspondence between the hierarchical annotation in Corbis and the hyponym generalization hierarchy in WordNet [14]

sets of synonyms and are referred to in WordNet’s documentation as *synsets*. These concepts and lemmas are organized into a network using a set of semantic and lexical relationships.

The overlap between the thesaurus used to annotate images in the Corbis collection and the hierarchies in WordNet enables us to derive relationships between the keywords in Corbis and the concepts in WordNet with an accuracy of over 90% for the most frequently occurring keywords [14]. Figure 6 shows an example of the overlap between the annotations of an image of an Arabian horse and the generalization hierarchy in WordNet.

### 3.6.2 Semantic interest metrics

Our implementation of semantic-guided browsing is based on several of the search tactics identified by Bates that people use to redirect their search to more general or more specific concepts when they find too many or too few results [1].

Detailed semantic structions such as WordNet are often difficult for novices to use for a number of reasons, e.g., their specialized vocabularies, internal inconsistencies, incomplete structure, or level of granularity. Figure 7 shows a portion of the overlapping generalization and composition hierarchies for the concept “horse.” For example, a person looking for “foal” and “mare” might be

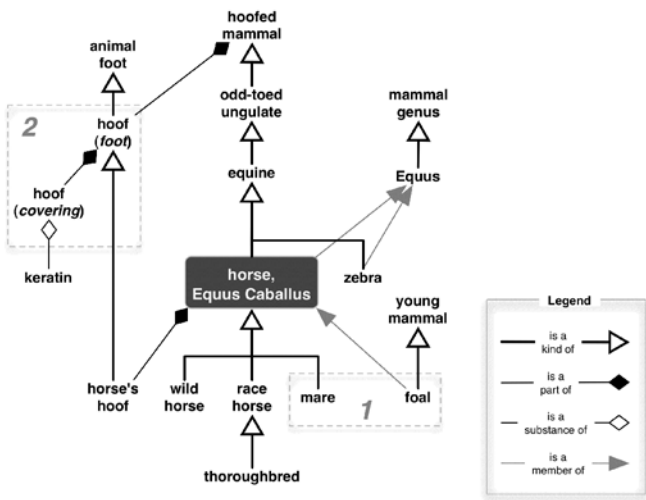


Fig. 7. A portion of WordNet showing the overlapping generalization and composition hierarchies

surprised that they are not related to “horse” in the same way (*region 1*), or a person looking for the substance of a horse’s hoof might be unable to find the concept where this information is encoded (*region 2*). To simplify the process of browsing using WordNet, we combine the three strategies to show the local neighborhood around the current focus.

We use a single complex composite metric to simultaneously support three basic search strategies: SUPER, SUB, and SIBLING. Each of these strategies is modeled as a directed search in WordNet along a particular type of relationship. Figure 8 shows examples of the semantic neighborhood that could be found using each of these strategies. For example, if a search for “horse” returned too many images, a person could use the SUB strategy to find a narrower set of more specific concepts, such as images of “wild horse” or “foal.” In order to apply these strategies without access to a semantic model such as WordNet, a person would have to know the existence of more general, more specific, or related concepts and add them to the query. Implementing these strategies using SFEV techniques allows a user to simply brush over a keyword and see the related concepts that exist in the image collection.

Figure 9 shows a data state model tracing the flow of information in a composite semantic metric where the user is focusing on a keyword. This model is similar to the models in Figs. 2 and 4, with the addition of a vertical region on the left that represents the subgraph of WordNet loaded into the browser. This subgraph contains all the concepts directly and indirectly related to the keywords loaded into the workspace.

When a user brushes over a keyword in the graphical model, the keyword is expanded to the different senses of the word. The composite metric expands the neighbor-

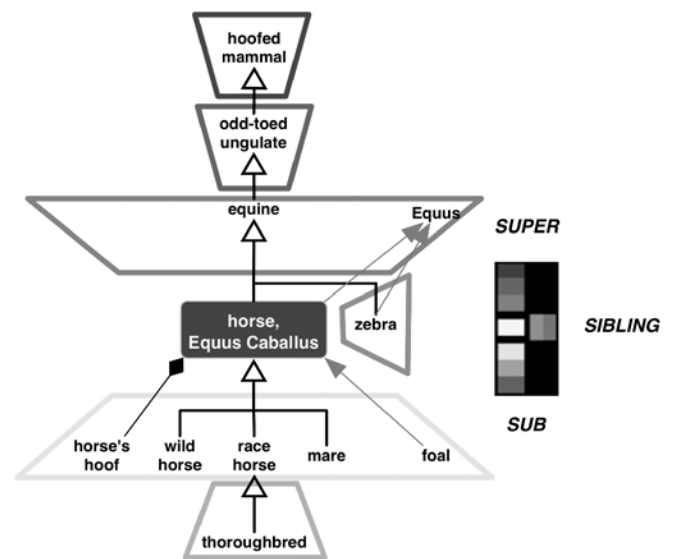


Fig. 8. A portion of the semantic neighborhood of the “horse” concept based on the SUPER, SUB, and SIBLING strategies

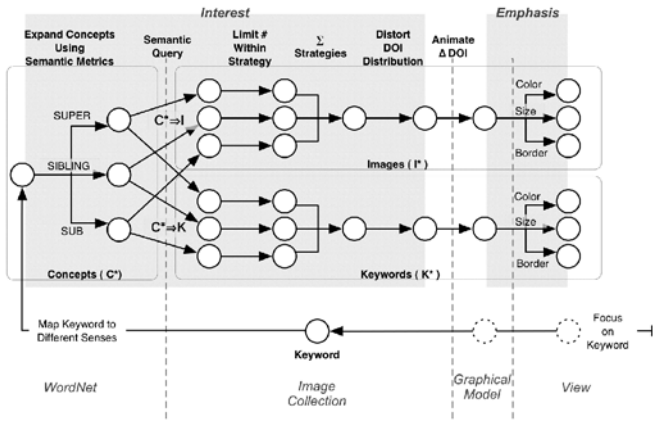


Fig. 9. Calculating interest with the SUPER, SUB, and SIBLING strategies using a composite interest metric

hood around each of these senses using the SUPER, SIBLING, and SUB strategies in parallel. The results of these metrics are a set of related concepts weighted by their distance from the focus. Each of these concepts is then mapped to a set of related images and keywords ( $C^* \Rightarrow I$ ). The set of images and keywords found within each strategy is assigned a DOI based on the path distance to the focus and the weight of the links between concepts, keywords, and images. The results from each strategy are limited and then aggregated so that the final view has a representative sample from each strategy. The remainder of the process is the same as described for Fig. 4. The keywords are assigned a different color to indicate the strategy that they result from, as shown in Fig. 10.

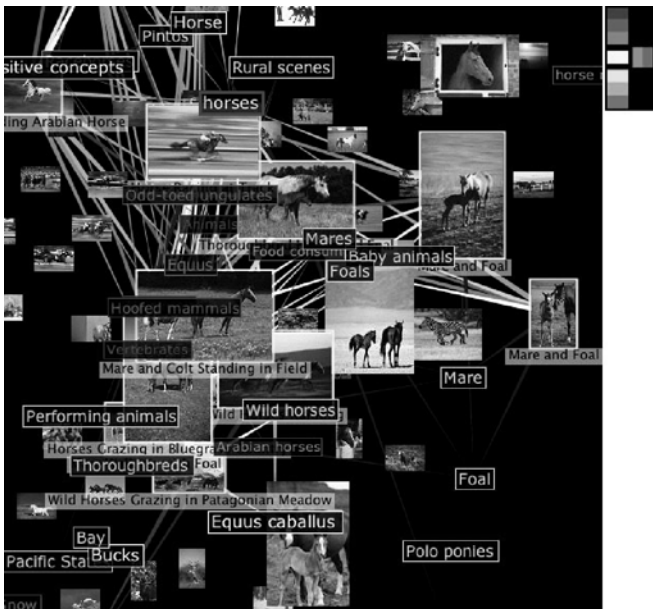


Fig. 10. Representing a sample of concepts most closely related to “horses” calculated with multiple strategies. The legend on the top right indicates that more general concepts are shown in red, more specific concepts are shown in cyan, and siblings are shown in magenta

### 3.7 Emphasis techniques

In the SFEV framework, emphasis techniques modify the visual representation of information to reveal changes in DOI. The goal of the emphasis techniques is to align the visual weight of objects with their importance in a particular context so that the most interesting objects are immediately apparent (i.e., “pop out”) and less interesting objects fade to the background. The relative contrast creates a visual ordering that allows a user to rapidly and opportunistically access contextual information. Although this approach is conceptually straightforward, in practice there are complex interactions between the visual scales used to encode interest, the distribution of DOI values, and the density of information in the display.

The most effective visual scale for presenting quantitative information, such as DOI, is position [7]. For this reason, position is almost always used by visualization techniques to encode the relationships between the primary data values. For example, we use position in this prototype to show global relationships between images and keywords using the spring layout. Graphical fisheye views often distort position and scale to increase the visibility and detail of information spatially near the focus. Because the interest metrics used by SFEVs are independent of spatial properties, distorting position to emphasize relative interest would be too disorienting to support rapid exploration. For this reason, we do not distort position in this prototype. Instead, we use a combination of size, color saturation, and density to encode DOI. These visual scales are less effective for precise quantitative comparisons of interest, but they are reasonably effective for showing categorical and ordinal information (e.g., membership in the set of interesting objects, and relative order).

We use different combinations of emphasis techniques to create different layers of information. Background information has high contrast and distortion so that only a small representative portion of the entire collection is shown in detail. Highlighted objects have higher minimum and maximum values to make them more visually apparent than most other information.

### 3.8 Interaction

Responsiveness is critical in maintaining a coherent visualization and allowing rapid opportunistic search; therefore, the interface uses brushing rather than selection to activate the focus of the fisheye view. This makes the interface highly sensitive to changes in the focus but also potentially disorienting and difficult to understand if too much information changes at the same time. The prototype uses animated transitions and multiple threads to immediately begin or immediately interrupt metrics when the focus changes.

A second critical element of SFEVs is controlling the visual complexity of transitions and final views. We used

hierarchical goal decomposition to analyze the sequence of actions and information required at different steps in the opportunistic task [4]. We noticed that in the process of analyzing the images in the view, users looked at information in a predictable sequence. Users would first rapidly scan over the images and captions in the collection. When they found an image that was interesting, they would pause to look at the keywords and compare them to the visual content of the image. The hierarchical goal decomposition allowed us to identify the knowledge required at each step of this process and limit the complexity of the view to the minimum information needed. Though subtle, the pause effectively separates the rapid course analysis of images and their captions from the more detailed analysis of correlating keywords and images.

#### 4 An experiment: comparing similarity-guided and semantic-guided search

Opportunistic search is an interactive process where an information seeker applies a variety of strategies in reaction to and incorporating the information they encounter while searching. In the context of exploring an unstructured annotated image collection, a person will browse over images and keywords, analyzing and synthesizing their content, and then using this information to guide them toward images that may be related in a variety of different ways. For example, in addition to looking for similar images, they may also look for more general, more specific, or alternative examples of a subject.

It is difficult to measure the effectiveness of an interactive search tool for the same reason that these tools are so essential; user needs are often complex and continuously change in reaction to the information encountered. An important component of interactive search is learning the vocabulary and semantic structure of the domain of interest. This domain knowledge will influence a user's ability to describe what they are looking for and to recognize and evaluate the usefulness of results. SFEVs are designed to interactively reveal information that is related to the user's continuously changing focus. In this section we describe the motivation, design, and results of a formal experiment comparing the effectiveness of two types of SFEVs for learning the domain knowledge that is essential for opportunistic search.

Standard metrics for evaluating interactive search interfaces are still evolving. For example, the interactive track at the 2003 Text REtrieval Conference (TREC) compared the effectiveness of interfaces for interactive Web-based "topic distillation" tasks. Users were asked to create a resource list of the most authoritative Web sites for a broad subject area. The interfaces were evaluated by comparing the resulting resource list using metrics such as the relevance and level of detail of each document selected and the coverage and repetition of the entire result

list, which are similar to the classic *precision* and *recall* metrics for evaluating information retrieval algorithms.

An alternative metric suggested by Pirolli et al. is the *incidental knowledge* learned by users during the search process [22]. In an experimental comparison of a similarity-based interface with the cluster-based Scatter/Gather interface, Pirolli et al. asked participants to draw concept maps of a document collection after they completed a search task. They found that the concept maps of participants using the Scatter/Gather interface were significantly broader and more detailed than those from the ranked-list interface. Based on these findings, they suggested that the Scatter/Gather interface might be particularly useful for *sensemaking* in unstructured collections of documents where the user is unfamiliar with the domain. Other researchers have proposed that information seekers mentally construct informal schemas as part of their *sensemaking* process to translate their domain knowledge into the vocabulary of the collection and to guide their search [25].

The TREC and Scatter/Gather experiments both *indirectly* evaluated the effectiveness of the interfaces for discovering the structure and diversity of a subject in an unstructured document collection. We initially considered designing a similar "incidental knowledge" experiment to evaluate the SFEV interactive visualization technique. The difficulty we encountered with designing a similar experiment was how to measure the discovery of domain knowledge. In pretests, we found that attention to peripheral information as well as the skill of developing concept maps varied greatly between participants. We eventually decided to reduce this variability by making the discovery of domain knowledge the primary task. Instead of asking participants to create a concept map after their search, we provide them with a partially filled concept map and ask them to find images and concepts that will complete the map. This is a more structured version of the TREC "topic distillation" task.

Our prototype provides two alternative methods for discovering domain knowledge in a collection of images. The first technique guides navigation by keyword similarity at the document (i.e., image) level, and the second reveals semantic relationships at the keyword/concept level. The second interface is designed specifically to enable navigation over generalization and specialization concept hierarchies, which Bates identified as a common search strategy [1].

The goal of our experiment is to compare the effectiveness of these two interaction techniques in a complex search task. Our primary hypothesis is as follows:

H1: Users will be more effective in opportunistic search and sensemaking tasks with an interface that reveals semantic relationships rather than keyword similarity.

The experiment is a within-subject design, with *interface* and *image collection* as the independent variables (Table 1). The two interfaces are identical except for the



interest metrics that the fisheye view is based on. The first interface uses content-based similarity metrics to calculate the distance between images, and the second uses interest metrics that calculate the distance between concepts. We carefully constructed two separate image collections of equal size and similar subject diversity.

Each participant completed one complex search task with each of the interfaces. We used two different image collections to avoid learning effects and counterbalanced combinations of variables between subjects to avoid order effects. A summary of the pattern of combinations for the first four participants is shown in Table 2 and was repeated to accommodate additional participants. In the following sections, we describe the design of this experiment in greater detail, followed by a discussion of results.

#### 4.1 The task

The interactive task for the experiment is complex and involves both opportunistically searching over the collection and making sense of the information encountered. We created two task scenarios that are identical except for the topic and are based on examples from the TREC 2003 Interactive Track. For the topics, we selected two different animal hierarchies because they are relatively easy for participants to understand, they are well-defined in both the image collection annotations and WordNet, and they are easy to recognize visually.

In each scenario, participants were told they were assistants for a biology course and were asked to find images for a Web site to accompany lectures on the subject “hoofed mammals” or “sea mammals.” The script for the “sea mammals” task is as follows:

*We are assistants in charge of a Web site for a biology course that will be taught next semester for 16-*

*year-olds. The professor is going to prepare a lesson on whales and other mammals that live in the ocean and would like us to find images to put on the Web site. The professor quickly sketched the basic organization of the Web site, but we need to fill in the details so he can finalize the course material according to the images available.*

After reading the scenario, participants were asked to complete a partially filled hierarchy by exploring a collection of relevant images to find keywords for each node and representative images for each leaf. Participants always began their search from a single visible image representing a concept at the deepest level in the hierarchy, for example “Beluga” in the “Sea mammals” taxonomy shown in Fig. 11. The taxonomy was partially filled with a representative concept from each level (shown in bold) and divided into regions that are progressively more distant from the starting image. To discover concepts that will complete the taxonomy, participants navigate over the relationships between images and keywords in the collection. The actual concept hierarchy is much more complex than the taxonomy shown above, and the images are related in diverse ways, as explained in the following section.

#### 4.2 The image collections

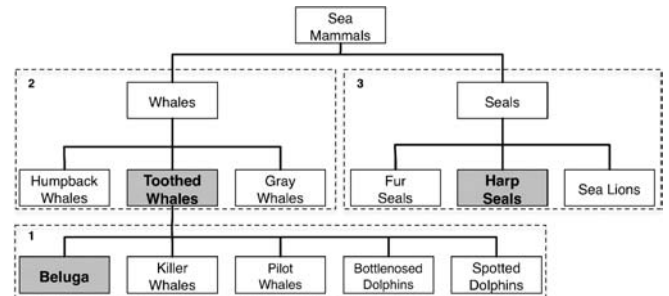
We carefully selected 100 images to represent the concepts in the “hoofed mammals” and “sea mammals” taxonomies. As described earlier in Sect. 3.2, most images contain a list of keywords that are a path to a specific concept in a classification hierarchy. Corbis added these lists to the images by expanding concepts using their proprietary thesaurus. For example, an image with the keyword “killer whales” will often also include the keywords “toothed whales”, “whales”, and “sea mammals.” These keywords may be mixed with many others describing other subjects in the image, but close inspection would allow a user to recognize or learn the existence of the more general concepts. Bates identified this type of navigation as the SUPER search tactic commonly used by information retrieval experts when there are too few results to a query [1].

**Table 1.** Independent variables

Variable	Instance	Description
1. Interface	<i>S</i>	Semantic distance
	<i>C</i>	Content similarity
2. Image collection	<i>Horses</i>	“Hoofed mammals”
	<i>Whales</i>	“Sea mammals”

**Table 2.** Within-subject experimental design. Each participant completed two tasks with a different combination of *interface type* (S,C) and *image collection* (H,W)

Participant	Task 1	Task 2
1	<i>S · H</i>	<i>C · W</i>
2	<i>S · W</i>	<i>C · H</i>
3	<i>C · W</i>	<i>S · H</i>
4	<i>C · H</i>	<i>S · W</i>



**Fig. 11.** The template for the “sea mammals” scenario. The nodes with the *gray background* were given, and participants were asked to find keywords and images to fill the remaining nodes

Although in most cases the concept hierarchy in the Corbis thesaurus closely matches the WordNet is-a-kind-of hierarchy, we had to correct several image annotations where there were significant conflicts between the WordNet and Corbis vocabulary and structure that would affect the experiment. First, we corrected differences in vocabulary for keywords that were required for navigating in the hierarchy. For example, Corbis uses the term “sea mammals” while WordNet uses “aquatic mammals.” In this case, we replaced the inexact match

$$k_i \Rightarrow \{l_1, l_2\}, \quad (2)$$

where the keyword  $k_i = \{\text{“sea mammals”}\}$  is mapped to the two lemmas  $l_1 = \{\text{“sea”}\}$  and  $l_2 = \{\text{“mammal”}\}$ , with the exact match

$$k_i \rightarrow \{l_1\}, \quad (3)$$

where  $k_i = \{\text{“sea mammals”}\}$  and  $l_1 = \{\text{“aquatic mammal”}\}$ .

Second, we corrected a structural conflict between Corbis and WordNet for the position of the “gray whales” concept. Corbis incorrectly places “gray whales” as a subcategory of “toothed whales” when they are actually “baleen whales.” Because of this error in the Corbis thesaurus, most images annotated with the keyword “gray whales” are also annotated with the keyword “toothed whales.” We repaired this error by replacing the link for “toothed whales” with a link for “baleen whales” in the affected images.

### 4.3 The interfaces

The two interfaces were identical in appearance and general functionality. At the beginning of the experiment, only the starting image was shown in detail; the rest of the images were represented as gray rectangles, and the keywords were not displayed. Both interfaces support spatial navigation using panning and zooming, and structural navigation by brushing over images and keywords.

The main difference between the two interfaces is the interest metric that they use for browsing over the collection. One interface uses content-based similarity metrics to find the ten images or keywords that are most similar to the focus. The second interface uses a combination of metrics based on semantic distance in WordNet and reveals a sample of the different senses of a keyword and a limited number of related concepts.

It is important to note that the hierarchical information necessary to complete the task is accessible in both interfaces. Many of the images are directly annotated with a set of keywords that corresponds to the path from the most specific concept present in the image up to a more general superconcept. The difference is that the relationships between the terms in the hierarchy are only explicitly shown in the semantic-based interface. When a participant does not find or recog-

nize related concepts using the content-based fisheye view, they cannot be certain of whether the annotations are missing or whether the relationship does not exist.

Participants controlled navigation indirectly by telling an expert what images and keywords they would like to see, but they did not physically interact with the interface. We noticed during trial studies that there was a very large variation in how quickly different users became comfortable controlling the interface, especially when the interface was slow to respond or reacted in unexpected ways. The goal of the experiment was to test how the interface supports semantic navigation, so we decided to use an expert “driver” to avoid keystroke-level problems. This also removed another dimension of variability between participants.

### 4.4 Experimental results

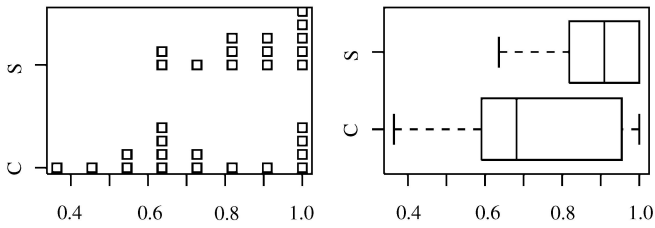
The main hypothesis of our experiment was that participants would be more effective at a complex directed search task with the interface that reveals *semantic* relationships within the collection rather than the interface that reveals *similarity*. We collected a range of quantitative and qualitative data during the experiment from the concept hierarchies that participants completed, their log files, and their questionnaires. In this section, we use the results of the experiment to compare the effectiveness of the two interfaces by analyzing user performance, browsing behavior, and satisfaction.

#### 4.4.1 Differences in performance

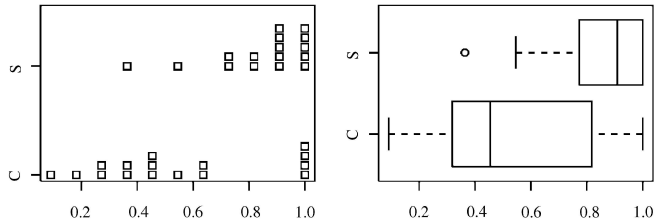
We evaluated the performance of the interfaces based on the accuracy of the participants’ concept hierarchies and their confidence. We also compared the time participants took to complete the hierarchies, although they were not under any pressure to finish quickly.

Summaries of the results comparing the interfaces for recall, accuracy, and confidence are shown in the sets of strip charts and box plots in Figs. 12–14. The upper set of data in each chart is for the semantic interface (S), and the lower set is for the content-similarity interface (C). Under each figure, we give the mean ( $\bar{X}$ ), median ( $M$ ), number of participants ( $n$ ), and probability of significance ( $p$ ). Using an ANOVA test, we found significant differences between the *interfaces* for each of these measures. We did not find any significant difference between image collections or the order in which participants used the interfaces for these results.

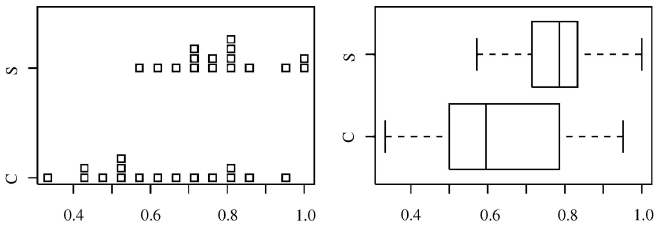
The results from this analysis suggest that the semantic interface was significantly more effective than the similarity interface for this complex search task. Participants using the semantic interface (S) found significantly more concepts ( $p < 0.02$ , Fig. 12), significantly more of the concepts they found were correct ( $p < 0.01$ , Fig. 13), and they were significantly more confident of the concepts they



**Fig. 12.** Percent of total concepts found (S:  $\bar{X} = 0.89$ ,  $M = 0.91$ ; C:  $\bar{X} = 0.73$ ,  $M = 0.68$ ;  $n = 16$ ,  $p = 0.02$ )



**Fig. 13.** Percent of concepts found that were correct (S:  $\bar{X} = 0.85$ ,  $M = 0.91$ ; C:  $\bar{X} = 0.55$ ,  $M = 0.45$ ;  $n = 16$ ,  $p < 0.01$ )

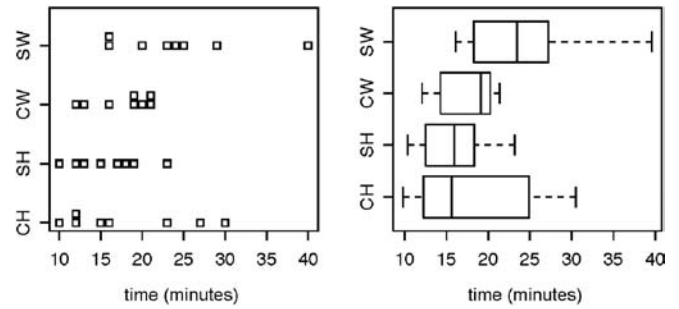


**Fig. 14.** Percent confidence that concepts found were correct (S:  $\bar{X} = 0.78$ ,  $M = 0.78$ ; C:  $\bar{X} = 0.63$ ,  $M = 0.6$ ;  $n = 16$ ,  $p < 0.01$ )

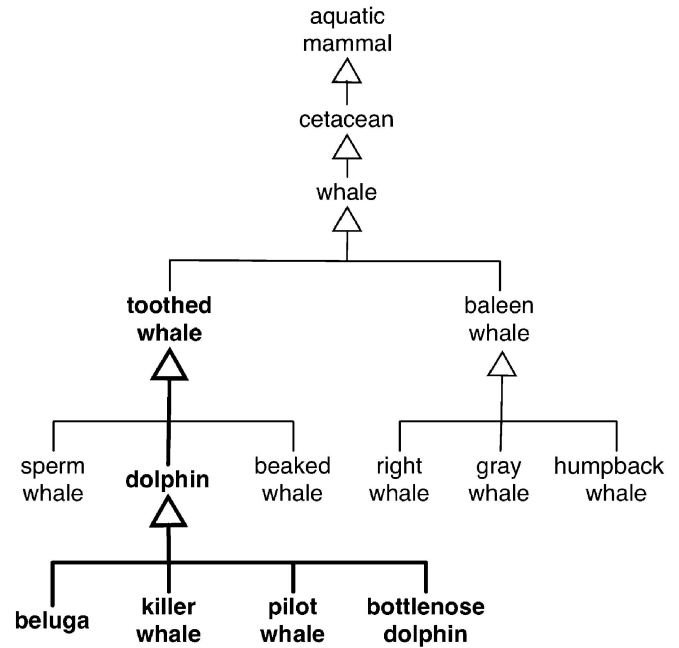
found ( $p < 0.01$ , Fig. 14) than with the content-similarity interface (C).

In our analysis of the time participants took to complete each task by *interface*, the ANOVA test did find significant differences between *image collections*. A comparison of the interfaces within each image collection found no statistical difference in total time for the Horses collection but indicated that participants using the semantic interface took significantly longer to complete the hierarchy in the Whales collection ( $n = 8$ ,  $p < 0.05$ ). Figure 15 shows the detailed results. This suggests that the Whales collection was more difficult to browse over using WordNet.

One likely contributor to this difficulty was a conflict between the general knowledge of the participants and the formal classification in WordNet. The most obvious example of this occurred in the relationship between *dolphins* and *whales*. In the biological classification that Corbis and WordNet share, dolphins are classified as a type of toothed whale, and beluga whales, killer whales, and pilot whales are classified as subtypes of dolphin (Fig. 16). However, most of the participants were hesitant to accept this classification. The



**Fig. 15.** Total time to complete the task, grouped by *interface* (S,C) and *image collection* (W,H). The difference between SW and CW is significant (SW:  $\bar{X} = 24$ ,  $M = 24$ ; CW:  $\bar{X} = 18$ ,  $M = 19$ ;  $n = 8$ ,  $p = 0.034$ )



**Fig. 16.** A portion of WordNet showing the relationship between whales and dolphins. Almost every user was confused by the classification of “killer whales” as a type of dolphin

longer time that participants took with the semantic interface could reflect the difficulty they had in resolving this conflict using the information they had available.

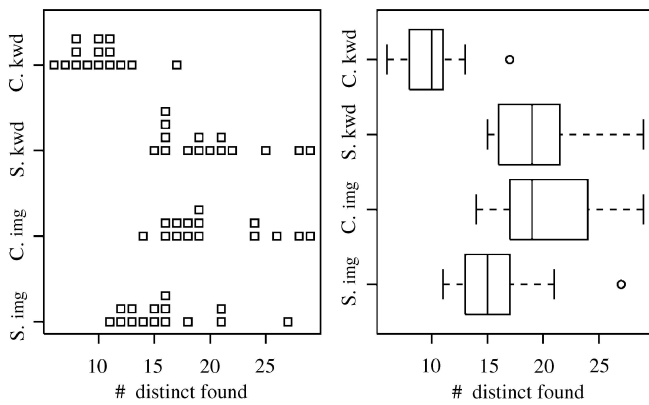
#### 4.4.2 Differences in browsing behavior

To analyze differences in browsing behavior, we extracted the number of distinct keywords and images visited by each participant as they were searching for concepts, as well as the rate at which they discovered new information. We found that participants using the semantic interface browsed to nearly twice as many distinct keywords as they did with the similarity interface but only slightly fewer images (Fig. 17). Figure 18 shows the number of distinct images and keywords browsed for each concept

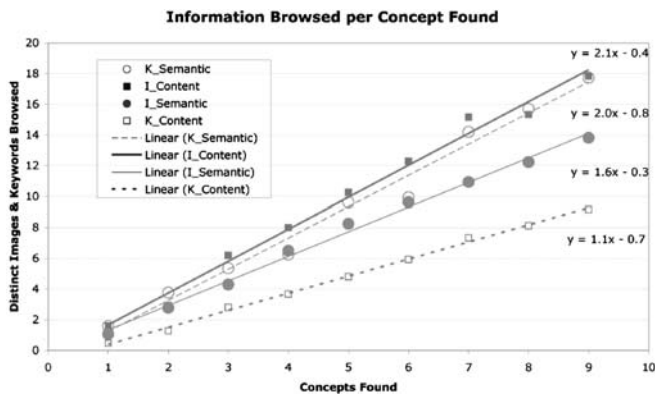
found. Both figures indicate that participants browse primarily with images in the content similarity interface and primarily with keywords in the semantic interface.

There are two likely reasons for this difference. First, each interface encourages the user to focus on different types of information in a collection. The *semantic* interface emphasizes the hierarchical relationships between concepts, so users will naturally focus on the keywords. On the other hand, the *content-similarity* interface emphasizes images that are similarly annotated, allowing users to navigate directly between images without selecting a keyword. Users select a keyword only when they want to limit their search to a particular feature. Second, the interfaces show the same number of images, but not the same number of keywords. The semantic interface may show 10 keywords in each of the SUPER, SUB, and SIBLING hierarchies, while the similarity interface shows only the ten most important keywords for an image. Participants will therefore see and have the opportunity to visit many more keywords with the semantic interface.

Close inspection of the keyword data for the semantic interface in Fig. 18 (the hollow circles labeled



**Fig. 17.** Distinct keywords and images browsed with each interface. Participants browsed significantly more keywords with the semantic interface (C.kwd:  $\bar{X} = 10$ ,  $M = 10$ ; S.kwd:  $\bar{X} = 20$ ,  $M = 19$ ;  $n = 16$ ,  $p < 0.001$ ), and significantly more images with the content similarity interface (C.img:  $\bar{X} = 20$ ,  $M = 19$ ; S.img:  $\bar{X} = 16$ ,  $M = 15$ ;  $n = 16$ ,  $p = 0.015$ )



**Fig. 18.** Average number of distinct images and keywords browsed per concept found with each interface

*K\_Semantic*) reveals three distinct groups that roughly correspond to the different regions in the task concept hierarchy (e.g., four concepts in the first region). These steps could correspond to the difference between searching for subconcepts and searching for siblings as users move progressively further from the start image. We would need to conduct additional experiments to further characterize these differences.

#### 4.4.3 Questionnaires

The exit questionnaires asked participants to compare the interfaces along a number of dimensions using a 7-point scale and to give comments about what they liked and disliked about the systems. The questions comparing the two interfaces followed the pattern of “*Which system is better for ...*”. Participants responded using a 7-point scale (1 = *Content-Similarity*, 4 = *Neither*, 7 = *Semantic*). In the actual questionnaires, the position of the interfaces at the ends of the scales changed to reflect the order in which the participant used the interfaces.

The first question asked participants how different they found the two interfaces. Visually, the two interfaces are nearly identical. However, during interaction the similarity and semantic-interest metrics reveal very different types of information. Participants found the interfaces to be very different (*Median* = 6).

Questions 3 and 4 asked which interface was “*easier to learn to use*” and “*easier to use.*” We expected participants to favor the *content-similarity* interface for ease of use because the *semantic* interface is visually and cognitively more complex: it presents more keywords at one time and superimposes multiple types of relationships that require a legend to decode. However, participants rated the semantic interface *easier* in both cases (Q3:  $M = 6$ ; Q4:  $M = 6$ ). One interpretation of these results is that the *semantic* interface is easier for the type of search that participants were asked to perform. This interpretation is also implied by several comments that identify general “ease of use” as a strength of the similarity interface:

(P1): “*Easier for a beginner. You don’t have to think, just point and click.*”

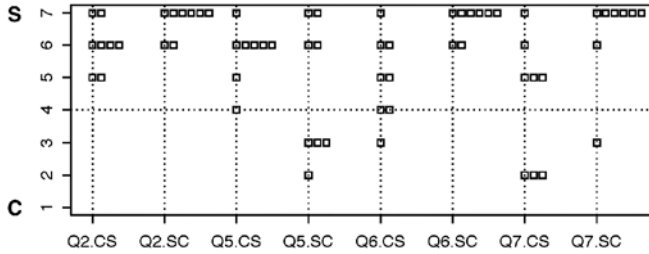
(P14): “*Less complicated. Simple.*”

(P16): “*Simple, accessible, instantly understood relationships.*”

The remaining questions compared the effectiveness of the interfaces for the task scenario. Using an ANOVA test, we found a significant difference in responses for several questions based on which interface the participants used first. Table 3 we summarize the responses by interface order, and Fig. 19 shows an overview of the distributions. The results suggest that participants that used the semantic interface first tended to prefer it more strongly for learning about the topic (Q2,  $p < 0.05$ ), finding different kinds of images (Q6,  $p < 0.01$ ),

**Table 3.** Median response from exit questionnaire comparing interfaces, separated by interface order

Question	CS	SC	p
Q2. <i>Learning about topic?</i>	6	7	$p < .05$
Q5. <i>Finding same kind of image?</i>	6	5	
Q6. <i>Finding different kinds of images?</i>	5	7	$p < .01$
Q7. <i>See more images?</i>	5	7	$p < .05$

**Fig. 19.** Responses to questions 2, 5, 6, and 7 of exit questionnaire, grouped by the order participants used the interfaces

and seeing more of the images in the collection (Q7,  $p < 0.05$ ). One possible explanation for this order effect is that participants develop a strategy for finding information using the first interface and then attempt to continue using the same strategy in the second interface. Transitioning from the guided exploration possible with the semantic interface to the unstructured exploration of the similarity interface is apparently more frustrating than vice versa.

Participants strongly preferred the semantic interface for finding both similar (Q5) and different kinds of images (Q6). We expected that the semantic interface would be more effective for finding different images because it is specifically designed to provide a sample of the diversity of related concepts. However, we did not expect that the semantic interface would be more effective for finding similar images than the similarity interface. A possible interpretation of this result is that participants understood “the same kind of image” to mean within a single concept rather than a combination of features at the document or image level. However, their comments about the similarity interface suggest that they were often frustrated by exactly this document-level type of similarity:

(P9): “Images were too similar.”

(P8): “It felt like I was turning in circles [of the same type of image].”

(P1): “Difficult to get out of an image topic. For example, it was difficult to get out of Orca because I was stuck in Dolphin.”

(P6): “Must find the most uncommon feature of an image to find other images.”

Several participants suggested that they would like to be able to refine and control the similarity metric:

(P16): “Search is often an intersection of concepts, so I want to be able to select more than one.”

Participants also preferred the semantic interface for learning about the topic structure (Q2) and exploring more of the images in the collection (Q7). Participants commented that the semantic interface was easier for discovering and verifying their understanding of the conceptual relationships between keywords and better for directed navigation.

We were surprised to find that the additional complexity of the semantic interface was not viewed as a problem, and that, although the interface required more concentration to use, the participants generally appreciated the additional information. On the other hand, opinions were mixed about the complexity of the similarity interface: some participants suggested that the interface should show more images, and others commented that there were already too many. Increasing the user’s control of the complexity is an important area for future research with this type of interaction technique.

Users unanimously appreciated the hierarchical structure visible in the semantic interface. They found this made it easier to discover and verify their understanding of the conceptual relationships between the keywords and was useful for structuring their navigation. Several participants said the semantic interface was instructive and would be useful as a tool for learning about a collection.

The most frequent weakness cited of the semantic interface was that it was difficult to perceive distance with saturation. Several participants also suggested that they would like to have automated commands to navigate along hierarchies and through history (e.g., upwards and backwards) and that the interface should show the relationships between objects more clearly. Developing more effective visual cues and interaction for SFEV techniques is an important area of future research.

#### 4.5 Discussion of results

The results of the experiment showed a strong advantage of the semantic-guided interface over the similarity-guided interface for opportunistic search and sensemaking tasks. Sensemaking, in particular, requires learning conceptual relationships such as the generalization and composition hierarchies shown by the semantic interface. Participants found significantly more concepts, made fewer errors, and had more confidence in their results with the semantic interface. The weaker performance of users with the similarity-guided interface suggests that making sense of the relationships between keywords without any visual cues is a difficult task. Furthermore, we found significant differences in browsing behavior between interfaces. With the semantic interface participants focused on the keywords to explore the collection, whereas the content-similarity interface allowed participants to focus primarily on images.

## 5 Conclusion

Opportunistic search and sensemaking in large information collections are highly interactive tasks that are poorly supported in current interfaces. These types of search activities require rapidly discovering, analyzing, and navigating between the relationships within information collections. A significant obstacle for users to effectively search over unstructured collections is their lack of domain knowledge, such as the vocabulary and semantic structure of the collection. We propose semantic fisheye views as an interactive visualization technique to support effective exploration over unstructured collections of information by guiding navigation using similarity and semantic relationships. Fisheye views reduce the visual complexity of displays by selectively emphasizing the most important information in a representation and deemphasizing or filtering less important information.

We have conducted an experiment comparing the effectiveness of similarity- and semantic-based fisheye views for complex search tasks. The results of this experimental evaluation suggests that the semantic-guided interface is significantly more effective for complex sensemaking tasks than the similarity-guided interface. These results strongly support semantic fisheye views as an effective interactive technique for opportunistic search and sensemaking tasks. An important implication of these results is that similarity-based search, which is based on algorithms used by common query and cluster techniques, is significantly less effective than semantic-guided search when *semantic diversity* is important.

The main contribution of this research is a characterization of some of the strengths and weaknesses of semantic fisheye views and of results that suggest that these are promising techniques for opportunistic search and complex sensemaking tasks.

*Acknowledgements.* This paper is based in large part on unpublished research in [12]. This research was funded by Grant 2000-066816 from the Swiss National Science Foundation and was supported by the generous permission of Corbis Corporation to directly access an electronic version of their annotated images. We also thank the helpful suggestions of the anonymous reviewers.

## References

- Bates MJ (1979) Information search tactics. *J Am Soc Inf Sci* 30:205–214
- Bates MJ (1989) The design of browsing and berrypicking techniques for the online search interface. *Online Rev* 13(5):407–424
- Bederson BB (2001) Photomesa: a zoomable image browser using quantum treemaps and bubblemaps. In: Proceedings of the 14th annual ACM symposium on user interface software and technology. ACM Press, New York, pp 71–80
- Bhavnani SK, Bates MJ (2002) Separating the knowledge layers: Cognitive analysis of search knowledge through hierarchical goal decomposition. In: Proceedings of the American Society for Information Science and Technology annual meeting, 39:204–213
- Casner SM (1991) A task-analytic approach to the automated design of graphic presentations. *ACM Trans Graph* 10(2): 111–151
- Chi EH-H, Pirolli P, Chen K, Pitkow J (2001) Using information scent to model user information needs and actions on the web. In: Proceedings of the ACM conference on human factors in computing systems, Seattle. ACM Press, New York, pp 490–497
- Cleveland WS (1985) The elements of graphing data. Wadsworth Advanced Books and Software, Monterey, CA
- Furnas GW (1986) Generalized fisheye views. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 16–23
- Furnas GW, Rauch SJ (1998) Considerations for information environments and the navique workspace. In: Proceedings of the 3rd ACM conference on digital libraries. ACM Press, New York, pp 79–88
- Hollink L, Schreiber G, Wielemaker J, Wielinga B (2003) Semantic annotation of image collections. In: Proceedings of the KCAP'03 workshop on knowledge capture and semantic annotation, Sanibel, FL. ACM Press, New York
- Ingwersen P (1994) Polyrepresentation of information needs and semantic entities: elements of a cognitive theory for information retrieval interaction. In: Proceedings of the 17th annual international ACM SIGIR conference on research and development in information retrieval. Springer, Berlin Heidelberg New York, pp 101–110
- Janecek P (2004) Interactive semantic fisheye views for information workspaces. PhD thesis, Swiss Federal Institute of Technology – Lausanne (EPFL)
- Janecek P, Pu P (2002) A framework for designing fisheye views to support multiple semantic contexts. In: International conference on advanced visual interfaces (AVI02), Trento, Italy. ACM Press, New York, pp 51–58
- Janecek P, Pu P (2003) Searching with semantics: an interactive visualization technique for exploring an annotated image collection. In: Meersman R, Tari Z (eds) OTM Workshops, Catania, Italy. Lecture notes in computer science, vol 2889. Springer, Berlin Heidelberg New York, pp 185–196
- Leung YK, Apperley MD (1994) A review and taxonomy of distortion-oriented presentation techniques. *ACM Trans Comput-Hum Interact* 1(2):126–160
- Marchionini G (1995) Information seeking in electronic environments. Cambridge University Press, Cambridge, UK
- Miller GA (1995) Wordnet: a lexical database for english. *Commun ACM* 38(11):39–41
- Motik B, Maedche A, Volz R (2002) A conceptual modeling approach for semantics-driven enterprise. In: Proceedings of the 1st international conference on ontologies, databases and application of semantics (ODBASE-2002). Lecture notes in artificial intelligence, vol 2519. Springer, Berlin Heidelberg New York
- Olston C, Chi EH-H (2003) Scenttrails: integrating browsing and searching on the web. *ACM Trans Comput-Hum Interact* 10(3):177–197
- Pecenovic Z, Do M, Vetterli M, Pu P (2000) Integrated browsing and searching of large image collections. In: Proceedings of the 4th international conference on visual information systems, Lyon, France. Springer, Berlin Heidelberg New York, pp 279–289
- Pirolli P (1997) Computational models of information scent-following in a very large browsable text collection. In: Proceedings of the SIGCHI conference on human factors in computing systems. PAPERS: Finding What You Want, vol 1. ACM, New York, pp 3–10
- Pirolli P, Schank P, Hearst M, Diehl C (1996) Scatter/gather browsing communicates the topic structure of a very large text collection. In: Proceedings of the SIGCHI conference on human factors in computing systems, Vancouver, BC, Canada. ACM Press, New York, pp 213–220
- Pu P, Janecek P (2003) Visual interfaces for opportunistic information seeking. In: Stephanidis C, Jacko J (eds) 10th international conference on human computer interaction (HCI '03), Crete, Greece. Erlbaum, Hillsdale, NJ, 2:1131–1135

24. Rodden K, Basalaj W, Sinclair D, Wood K (2001) Does organisation by similarity assist image browsing? In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM Press, New York, pp 190–197
25. Russell DM, Stefik MJ, Pirolli P, Card SK (1993) The cost structure of sensemaking. In: Proceedings of the SIGCHI conference on human factors in computing systems, Amsterdam. ACM Press, New York, pp 269–276
26. Sarkar M, Brown MH (1994) Graphical fisheye views. *Commun ACM* 37(12):73–84
27. Yee K-P, Swearingen K, Li K, Hearst M (2003) Faceted metadata for image search and browsing. In: Proceedings of the conference on human factors in computing systems. ACM Press, New York, pp 401–408
28. Zhang J, Norman DA (1994) Representations in distributed cognitive tasks. *Cogn Sci* 18:87–122