

Searching with Semantics: An Interactive Visualization Technique for Exploring an Annotated Image Collection

Paul Janecek and Pearl Pu

Human Computer Interaction Group,
Swiss Federal Institute of Technology, Lausanne (EPFL)
CH-1015 Ecublens, Switzerland
{paul.janecek, pearl.pu}@epfl.ch
<http://hci.epfl.ch/>

Abstract. Effective opportunistic search in visual interfaces for large information collections, such as documents, web pages, and annotated images, often requires domain knowledge that the information seeker does not have. Relevant semantic information, such as metadata and especially ontologies, are valuable resources for helping the user learn this domain knowledge. In this paper we describe an interactive “focus + context” visualization technique for integrating this semantic information into the representation of the information collection to support effective opportunistic search. We have implemented a prototype using these techniques that integrates WordNet, a general lexical ontology, and a large collection of professionally annotated images. As the user explores the collection, the prototype dynamically changes the visual emphasis and detail of the images and keywords to reflect the relevant semantic relationships. This interaction allows the user to rapidly learn and use the domain knowledge required for effective opportunistic search.

1 Introduction

Exploring a large, unfamiliar collection of information using queries is like ordering at an exotic restaurant without the benefit of a menu. It is very difficult to access or discover the rich selection just outside our limited vocabulary and knowledge, or to develop a coherent model of the global structure. Visual information retrieval interfaces (VIRIs) aid in the exploration task by representing the collection and its structure, as well as maintaining the users history of queries and results.

We are investigating how to extend VIRIs to more directly support the variety of strategies that people use for opportunistic search. Many of these strategies require access to domain knowledge that is not modeled in the interface. Our approach is to integrate semantic information, such as ontologies, into the interface to enable them to explore the information collection at a semantic level. In this paper we describe opportunistic search in greater detail, and discuss the

design of visual interfaces for supporting opportunistic search, especially “focus + context” techniques. We then give an overview of a prototype we have developed for exploring a large annotated image collection using the semantic concepts modeled in WordNet.

2 Opportunistic Search

Unlike the single-shot query where a person knows exactly what they are looking for, opportunistic search is more like research. The “information seeker” must learn a set of new concepts and relationships so they can efficiently navigate in an unfamiliar conceptual space, and evaluate the relevance of the information they find. Bates described this process as “berry-picking” because the query is satisfied by the information gathered along the search process (such as references, influential authors, and well-respected resources), rather than a single set of results (Fig. 1 shows an example)[1].

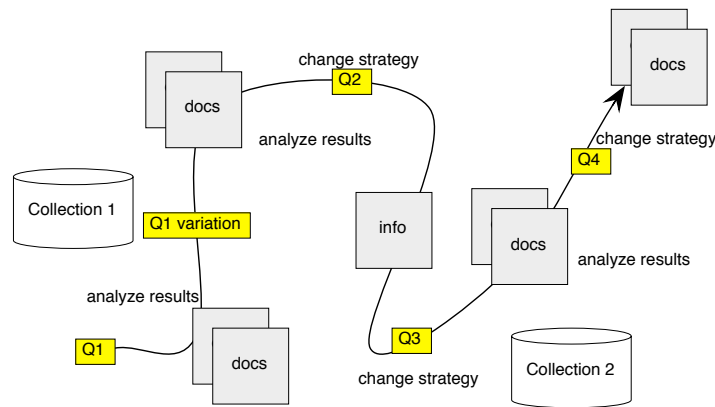


Fig. 1. Opportunistic search is a process of queries and exploration that is dependent on both the information encountered as well as the search strategies used

Bates [2] identified four general types of search tactics that “experts” use to find information:

- *monitoring tactics*, which manage the overall search process (e.g., record promising search paths that have not yet been explored);
- *file structure tactics*, which are techniques for effectively navigating through different information structures (e.g., reviewing all options before committing);
- *search formulation tactics*, which guide the design of the query (e.g., expand a query by adding synonyms); and
- *term tactics*, which guide the choice of words used in a query (e.g., refine a query by selecting a more general or more specific term).

Bates remarks that most interfaces for information retrieval do not directly support these higher-level strategies [3]. Recent advances in VIRIs lay the foundations for enabling these tactics. For example, Furnas' desiderata for the design of VIRIs [4] suggests visual reification of the "conceptual" aspects of search, such as history, and that information should always make its internal structure accessible for exploration. A VIRI enabled with these properties would facilitate the *monitoring* and *file structure tactics* mentioned above.

However, a significant obstacle that remains is that many of these tactics require access to domain knowledge that is not available in the interface. For example, to apply *search formulation* and *term tactics* the user must know the existence of more general, more specific, or related terms (i.e., the type of knowledge increasingly accessible in metadata and ontologies). Our approach to solving this problem is to integrate this semantic information directly into both the visual and interactive aspects of the interface to facilitate these strategies for opportunistic search.

3 Semantic Visualization and Interaction Techniques

A general problem in visual representations of large collections of information is managing the amount of information that is visible at any given time. Fish-eye view visualization techniques, also known as *focus + context* techniques, are a well-known approach to solving this problem by maintaining a balance between local detail and global structure. We have identified four important general properties that a VIRI should have to support opportunistic search using these techniques [5]:

- *Focus + context*: The balance between the local detail around the users focus and its context in the global structure must allow the user to perceive potential directions for future search. The focus may be a wide variety of objects in the interface: a query, a set of results, an object or one of its attributes, or perhaps an entire collection of objects. The context is the relationship between the focus and the rest of the information in the collection.
- *Dynamic selection of focus + context*: The user must be able to easily shift both their focus and context to explore different search strategies, (for example, across different domains, levels of granularity, or collections).
- *Multiple foci and multiple contexts*: Opportunistic search is multi-directional. The user must be able to compare the results of different foci and contexts to detect patterns.
- *Flexible modeling of semantics*: As a user explores an information collection, they will gather "interesting" information with undefined or incomplete semantics. The user must be able to externalize their understanding of the information they encounter during the search process in a "light-weight" way.

Our prototype, described in Sect. 4, is being developed to examine the interaction of these properties in opportunistic search.

3.1 Semantic Fisheye Views

Furnas originally defined fisheye views with the following equation [6]:

$$DOI(x|fp = y) = API(x) - dist(x, y). \quad (1)$$

The Degree of Interest (*DOI*) of each object, x , given a focus, fp , is the difference between the *a priori importance* (*API*) of x , and the distance between x and the focus. The visual representation of each object is modified interactively to reflect *DOI*. For example, Furnas implemented a fisheye view using the hierarchy of a structured programming language where *API* was the distance from the root, and *dist* was the path distance between nodes. As the user navigated through the code, lines of code with a *DOI* below a certain threshold were filtered resulting in a very compact display that balanced local detail and global context. However, equation 1 is general and has been redefined in many different data structures such as tables and graphs [7], [8].

In *graphical fisheye views*, the *API* and *dist* are defined spatially within the view or structurally within the model that is being visualized. In the case of opportunistic search, where there are potentially many different types of interrelated data collections, metadata and ontologies, there are also potentially many different types of distance metrics. Therefore, we have generalized equation 1 to include the possibility of balancing multiple weighted distance metrics as shown in equation 2 [9]. This equation adds a weight, w , to each distance function to allow a user to simultaneously explore multiple notions of relevance. *Semantic fisheye views* are based on this framework, with one or more distances defined semantically (e.g., within a related semantic model, such as an ontology), rather than spatially or structurally.

$$DOI_{context}(x|fp = y) = f(API(x), w_i, dist_i(x, y)) . \quad (2)$$

Once *DOI* is calculated, the visual representation is modified using one or more *emphasis algorithms* to allow the user to perceive the relative *DOI* of objects. For example, the size, color, grayscale, or level of detail may be modified, or information below a certain threshold may be filtered. In the next section we describe our prototype based on these techniques.

4 A Prototype for Exploring an Annotated Image Collection

We have implemented a VIRI prototype using *semantic fisheye views* to support opportunistic search in a very large collection of annotated images. Similar to NaviQue [4], the interface allows the user to search for images by defining a query or by browsing the images and their related keywords in a 2D space. The environment shows a history of previous searches and allows the user to gather interesting images and keywords.

Unlike NaviQue and other VIRIs, however, the prototype uses a related semantic model to support opportunistic search in two different ways. First, as the

user explores the collection the *semantic fisheye view* visually emphasizes keywords and images that are semantically related to the current focus. This serves as an explanation of the vocabulary of the collection, as well as a suggestion of potential directions for search that are more general, more specific, or similar to the current focus. Second, the semantic model aids in query disambiguation and, more importantly, query expansion to include related concepts that will return images in the semantic neighborhood of the original query.

In the following sections we introduce the functionality of the prototype with a scenario, describe how the prototype uses a semantic model and interactive visualization to support the *search formulation and term tactics* identified earlier, and a brief discussion of the architecture.

4.1 Searching with Semantics: A Scenario

A mother and daughter are looking for images of horses to use in a scrapbook they are making. They begin the search by typing in "horses", and over 500 images fill the space as shown on the top left of Fig. 2. Although it is easy to zoom in on any group of images, there are too many images and they decide to search for a specific kind of horse. Typing in "Arabian horses" they find the two images shown in the upper right figure. Placing the mouse over one of the images shows different words associated with the image, and placing the mouse over one of the words shows the related images as well as related concepts. When they place the mouse over "Arabian horses", as shown in the bottom left figure, they see the related concepts "Knights", "Bay", "Wild horses" and "Pintos". The daughter wonders what a Pinto is, and placing the mouse over the word she immediately sees images of Pinto horses and decides to include one in her scrapbook. Continuing in this way, they can find other types of horses or concepts they are unfamiliar with or hadn't considered.

In this scenario, the prototype supports opportunistic search by transparently suggesting new directions for their search. Like most VIRIs, the layout places related objects next to each other. The *semantic fisheye view* builds on this by emphasizing objects that are semantically related but not necessarily visually adjacent. The semantic relationships used in the scenario are the term tactics Bates identified [2]:

- SUPER: To move upward hierarchically to a broader term
- SUB: To move downward hierarchically to a more specific term
- RELATE (SIBLING): To move sideways hierarchically to a coordinate term

In the context of the semantic model we use in our prototype, WordNet, these tactics are transitions from a concept to neighboring concepts, as shown in Fig. 3.

We use a variety of visual emphasis techniques for displaying *DOI* dependant on the type of focus and the distance metrics used to define the context. Table 1 shows several examples of focus, context and emphasis techniques for Images (*I*), Keywords (*K*), and Concepts (*C*).

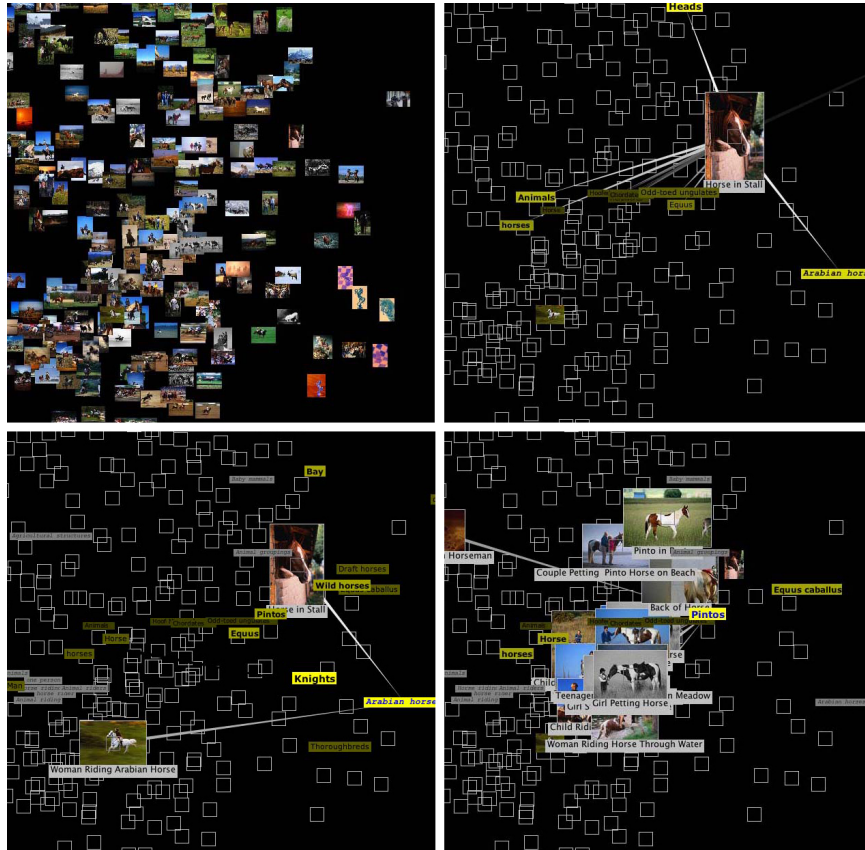


Fig. 2. An example session of browsing the semantic relationships between images

Table 1. Focus, Context, and Emphasis Techniques

Focus	Context	Emphasis
I	$I \rightarrow K^*$: term order	font size, background shading, edge grayscale
K	$K \rightarrow I^*$: cosine similarity measure	image size, edge grayscale
C	$C \rightarrow K^*$: semantic path distance	font size, background shading

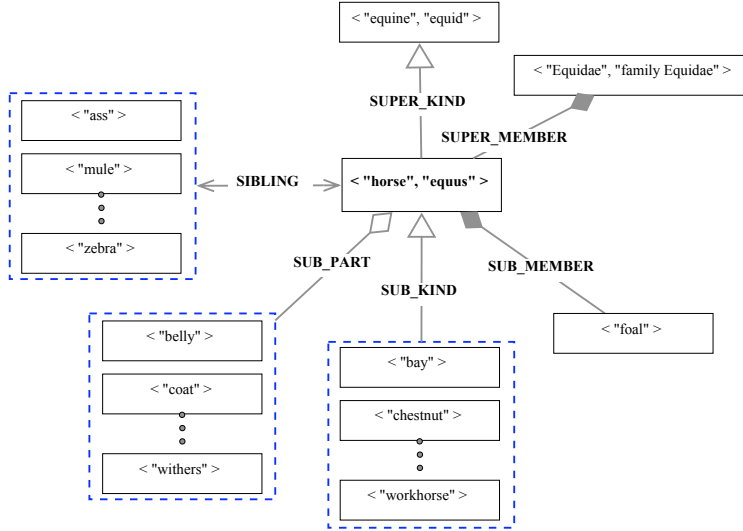


Fig. 3. The SUPER, SUB, and SIBLING tactics using WordNet

4.2 The Corbis Image Collection

On the left of Fig. 4 is a model of our annotated image collection, a set of 56,000 images from the Corbis collection. The figure shows the caption and keywords associated with an image of an Arabian horse. More formally, an image in the Corbis collection, I_i , is represented by an ordered set of one or more keywords, K_i^* , as shown in equation 3.

$$I_i \rightarrow K_i^* = \{k_1, \dots, k_n\}. \quad (3)$$

The images are professionally annotated with an average of over 20 keywords per image describing the objects, actions, and thematic elements of the images. Many of the annotations are hierarchical, meaning that an image of “Horses” will often also include the keyword “Animals” to distinguish it from horses that are not animals (such as “pommel horses”).

4.3 WordNet

WordNet is a large general ontology of the English language developed at Princeton [10]. As shown on the right of Fig. 4, each concept in WordNet is represented by a brief description and a *synset*, a set of synonymous words or lemmas. More formally, a concept in WordNet, C_j , is represented by a set of one or more synonymous lemmas, L_j^* , as shown in equation 4.

$$C_j \rightarrow L_j^* = \{l_1, \dots, l_n\}. \quad (4)$$

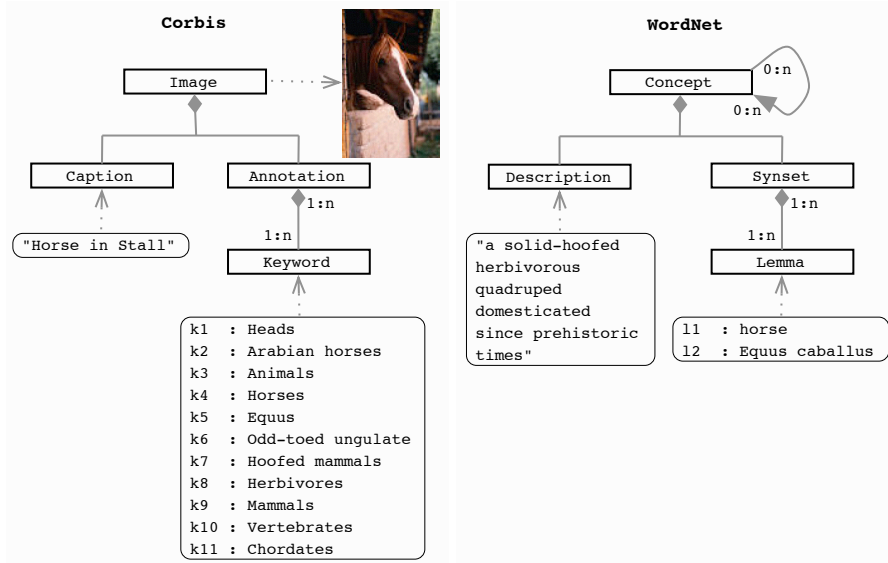


Fig. 4. Models of the Corbis annotated image collection and the WordNet lexical ontology. The concepts in WordNet are organized into a network structure based on a limited set of relationship types

The figure shows the description and lemmas associated with the “horse” type of animal. WordNet is large: there are approximately 140,000 lemmas linked to 110,000 concepts (the same word may have multiple meanings). The concepts are organized into a network using a limited set of semantic and lexical relationships, such as the *Is-A* (hyponym/ hypernym) and *Part-of* (meronym/ holonym) hierarchies for nouns.

4.4 Developing the Correspondence between Corbis and WordNet

In order to explore the image collection at a semantic level (e.g., browse the conceptual relationships between images), we created a mapping between Corbis and WordNet. In our current implementation, the correspondence is only a lexical mapping between the keywords in Corbis (K) and the lemmas in WordNet (L). As shown in Fig. 5, we were able to achieve a correspondence between K and L of over 90% for the 25% most frequent keywords using basic stemming rules (e.g., removing plurals and hyphenated terms). This correspondence drops to under 60% for the keywords that are only used once in the image collection. The majority of these matches are *exact*, for example the keywords “Horses” and “Horse” are both matched to the lemma “horse”. *Inexact* matches are those compound keywords where each subcomponent is exactly matched. For example, the keyword “four-wheeled carriage” is inexactly matched to the lemmas “four-wheeled” and “carriage”.

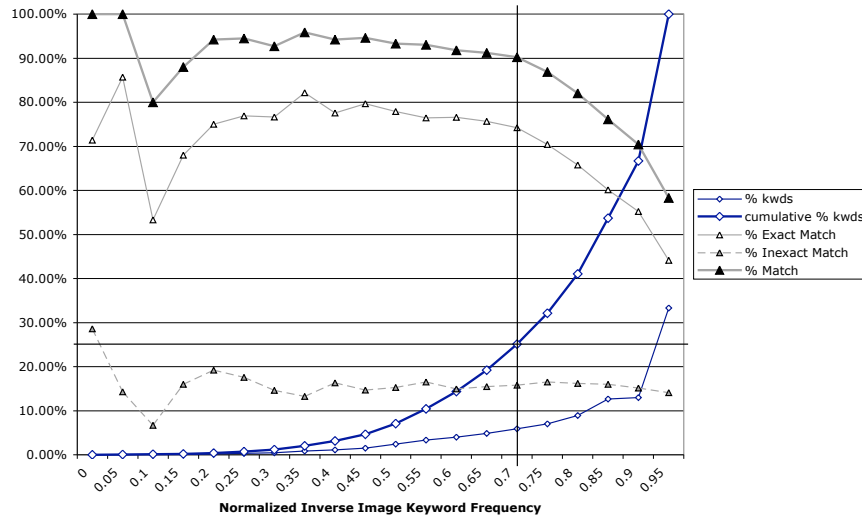


Fig. 5. The accuracy of the mapping between Corbis keywords and WordNet lemmas. The x axis is an inverse function of the frequency a keyword appears in the image collection. Keywords that appear most frequently have lower values on the x axis. The intersection of the lines $x=.725$, $y=25\%$ shows that the 25% most frequently used keywords are matched to lemmas with greater than 90% accuracy (75% exact and 15% inexact)

4.5 Using WordNet to Expand Queries

The imprecise correspondence between Corbis and WordNet introduces multiple levels of ambiguity when trying to search for images using concepts. In practice, however, the user is able to rapidly choose the examples that are most relevant to their search. Fig. 6 shows a comparison between finding images using keywords and using concepts. Both queries begin from text that must be matched to a keyword, $T \Rightarrow K$, or lemma, $T \Rightarrow L$ (*lexical disambiguation*). To use concepts, we must determine which sense of the lemma is correct, $L \Rightarrow C$ (*semantic disambiguation*), then how to represent this concept in keywords, $C \Rightarrow K$ (*semantic disambiguation*).

Fortunately, many of the images are hierarchically annotated (presumably with a tool like WordNet) to aid in semantic disambiguation. For example, Fig. 7 shows the overlap between the keywords associated with the image shown earlier (on the left in bold) and the corresponding concept in the *Is-A* hierarchy for an Arabian horse in WordNet (on the right in bold).

This semantic redundancy in the annotation enables a user to find images when they search using general terms. However, hierarchical annotation by itself does not provide a structure for exploring the collection. On the other hand, using a general semantic model, the user would be able to navigate along the relationships defined in the model, but not necessarily instantiated in the collection. Our prototype uses both together, allowing the user transparent access

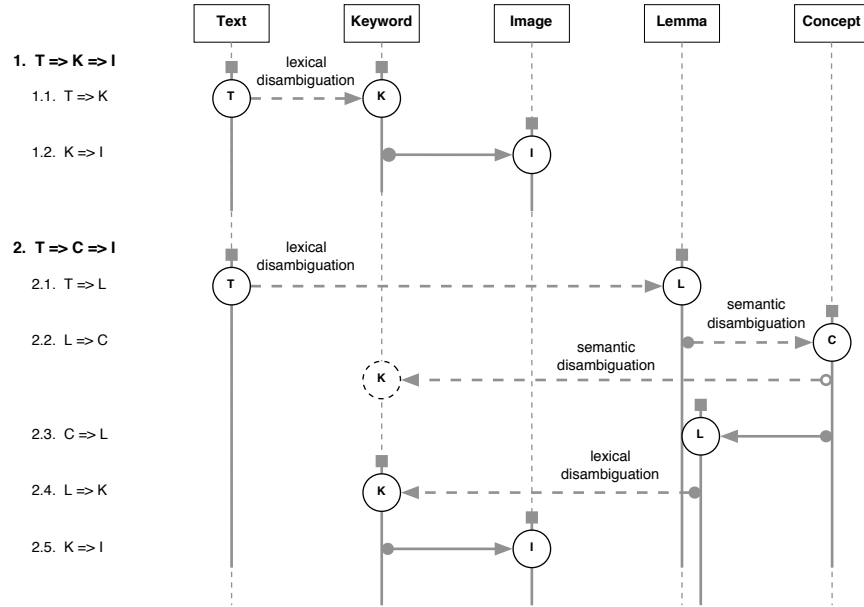


Fig. 6. A comparison of searching for images by keyword and searching by concept. Searching with concepts introduces the semantic ambiguity of selecting the correct sense of the lemma, $L \Rightarrow C$, and lexical ambiguity because of the incomplete match between lemmas and keywords, $L \Rightarrow K$


	Corbis	WordNet
	k1 : Heads	c1 : Arabian, Arab
	k2 : Arabian horses	c2 : saddle horse, ...
	k3 : Animals	c3 : horse, Equus caballus
	k4 : Horses	c4 : equine, equid
	k5 : Equus	c5 : odd-toed ungulate, ...
	k6 : Odd-toed ungulate	c6 : ungulate, hoofed animal
	k7 : Hoofed mammals	c7 : placental, ...
	k8 : Herbivores	c8 : mammal
	k9 : Mammals	c9 : vertebrate, ...
	k10 : Vertebrates	c10 : chordate
	k11 : Chordates	c11 : animal, ...

Fig. 7. The correspondence between the hierarchical annotation in Corbis, and the *Is-A* hierarchy in WordNet

to the structure of the model to navigate within the image collection. In the following section we define more formally how we create a semantic fisheye view using images, keywords and concepts.

4.6 Modeling “Focus + Context”

A semantic fisheye view is created by determining the degree of interest of each object in a collection with respect to the users current focus, as described earlier in equations 1 and 2. In our interface, we use the following general DOI equation:

$$DOI_{context}(x|fp = y) = API(x) - \sum_{i=1}^n w_i dist_i(x, y) . \quad (5)$$

The focus can be defined as a tuple composed of weighted vectors of keywords, images, lemmas, concepts, and previous foci:

$$f_i = \langle K_i^*, I_i^*, L_i^*, C_i^*, f_{n < i}^* \rangle . \quad (6)$$

In order to determine the distance between objects in different domains, we project the focus into the domain of interest. In addition to the unambiguous mappings from images to keywords, $I \rightarrow K^*$ (equation 4), and concepts to lemmas, $C \rightarrow L^*$ (equation 5), we define the following additional ambiguous transformations, where w is the probability of the match:

$$L_i \Rightarrow C_i^* = \{w_1 c_1, \dots, w_n c_n\} . \quad (7)$$

$$L_i \Rightarrow K_i^* = \{w_1 k_1, \dots, w_n k_n\} . \quad (8)$$

$$K_i \Rightarrow L_i^* = \{w_1 l_1, \dots, w_n l_n\} . \quad (9)$$

$$K_i \Rightarrow I_i^* = \{w_1 i_1, \dots, w_n i_n\} . \quad (10)$$

A fisheye view is created by determining the distance from the user’s focus to the related objects in the collection using these equations. The probabilities are then represented visually as described in Table 1, and shown in Fig. 2.

5 Conclusions and Future Work

Opportunistic search, exploration and semantic discovery in large collections of information can greatly benefit from interfaces that are able to integrate relevant semantic models with interactive visualization. In this type of environment, a perfect match between the collection and the semantic model is not necessary for effective opportunistic search. A user of this type of environment is at the early, exploratory phase of search where it is important to gain a broad overview of the concepts and vocabulary of a collection.

We have presented a prototype environment for exploring an annotated image collection based on *semantic fisheye views*. This prototype integrates WordNet by deriving a mapping between the keywords used to annotate images and the

lemmas associated with WordNet concepts. The user is then able to search for images based on the semantic relationships between concepts defined in WordNet. As the user changes their focus by navigating between images, keywords, and concepts, the *semantic fisheye view* modifies the representation of the objects in the collection to reflect the users *degree of interest*. The visual *emphasis techniques* reduce the complexity of the display, and the fluid interaction allows the user to discover new domain concepts and relationships, and suggest future directions for search.

References

1. Bates, M.J.: The design of browsing and berrypicking techniques for the online search interface. *Online Review*, **13** (1989) 407–424
2. Bates, M.J.: Information Search Tactics. *Journal of the American Society for Information Science*, **30** (1979) 205–214
3. Bates, M.J.: Where should the person stop and the information search interface start? *Information Processing and Management*, **26** (1990) 575–591
4. Furnas, G.W. and S.J. Rauch: Considerations for Information Environments and the NaviQue Workspace. In: *Proceedings of the 3rd ACM International Conference on Digital Libraries (DL '98)*. (1998) 79–88
5. Pu, P. and P. Janecek: Visual interfaces for opportunistic information seeking. In: Stephanidis, C., Jacko, J. (eds): *Proceedings of the 10th International Conference on Human Computer Interaction (HCI '03)*. Crete, Greece: Lawrence Erlbaum Associates (2003) 1131–1135
6. Furnas, G.W.: Generalized Fisheye Views. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '86)* . (1986) 16–23
7. Rao, R. and S.K. Card: The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus+Context Visualization for Tabular Information. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '94)* . (1994) 318–322
8. Sarkar, M. and M.H. Brown: Graphical Fisheye Views of Graphs. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '92)*. (1992) 83–91
9. Janecek, P. and P. Pu: A Framework for Designing Fisheye Views to Support Multiple Semantic Contexts. In: *Proceedings of the ACM International Conference on Advanced Visual Interfaces (AVI '02)*. ACM Press: Trento, Italy. (1992) 51–58
10. Miller, G.A.: WordNet: a lexical database for English. *Communications of the ACM*, **38** (1995) 39–41