Modeling Navigation in Degree-of-Interest Trees

Raluca Budiu (budiu@parc.com)
Peter Pirolli (pirolli@parc.com))
Palo Alto Research Center,
3333 Coyote Hill Dr., Palo Alto, CA 94304

Abstract

We present an ACT-R (Anderson & Lebiere, 1998) computational model of how people navigate in a degree-of-interest (DOI) tree. The model incorporates a visual salience function that determines which part of the display to attend to next. The salience function uses visual features of the display (e.g., distances) and semantic features of labels (e.g., information scent). The model was compared against data from participants and provided medium to strong fits to latencies and the number of nodes visited by the participants. The model shows that it is useful to distinguish between category-based versus similarity-based information scent. It also suggests that visual distance and scent may interact with one another, with scent playing a greater role at distances close to the current node in the visual focus.

Keywords: Degree-of-interest trees, information scent, information visualization, computational modeling, ACT-R.

Introduction and Motivation

One of the most prevalent tasks of contemporary society is information search and information browsing. According to *Pew Internet Survey* (2006), as of December 2006, out of the 141 million of Internet users in the US, 91% of them engage in information search activities (and they do it as often as they read email). Web pages often contain many links, and people are relatively good at finding their way through the information maze. Due to their limited information processing resources, people rarely perform an exhaustive search in the whole space of alternative paths; rather they rely on heuristics to reach their information goal. These heuristics involve both visual cues taken from the display and semantic cues offered by the words on the page.

In this paper, we study the nature of the search heuristics that people use when they interact with a complex, dynamic and information-rich display that changes at every click. Our work builds on two research areas: visual search and web navigation. There is a rich literature in psychology on visual search (see Wolfe, 1998 for a review), but most of it addresses simple visual displays and ignores the semantic content of the items on the screen. Since the emergence of the Web, several models of menu/web navigation have been proposed (Miller & Remington, 2004; Fu & Pirolli, in press; Blackmon, Kitajima, & Polson, 2005), but they deal with simple navigation tasks and predict only few navigation choices (e.g., given that the user and the model have clicked on the same sequence of choices, they strive to predict the next user choice, rather than the entire sequence leading to the solution).

The task that we study involves searching for a target in a particular search interface called degree-of-interest (DOI) tree (Card & Nation, 2002). The questions that we ask are: Does it matter where you place objects on the screen? What influences visual search most? Do people get distracted by irrelevant, but visually salient items? What is more important: the visual organization of the display or the words that are used? How do people decide what items are relevant? How can we estimate the relevance of words? Answers to this questions could have practical implication for interface designers, as well as for systems that leverage on user models to make relevant information more available to users.

We propose DOI-ACT, a cognitive model of navigation in DOI trees. The model is implemented in ACT-R (Anderson & Lebiere, 1998) and can make a successful sequence of navigation choices and reach the solution of the search problem in time (and number of clicks) comparable to humans.

Degree of Interest (DOI) Trees

Degree of interest trees (Card & Nation, 2002) are a form of focus+context visualization (Card, Mackinlay, & Schneiderman, 1999), intended to help with displaying large amounts of data. They compromise between a ”panoramic” and a ”zoom” view of the information, by using degree-of-interest calculations to decide which nodes should be visible. Figure 1 shows a degree-of-interest tree that represents an ontological hierarchy of concepts. The ontology was taken from the Great CHI’97 Browse-Off (Mullet, Fry, & Schiano, 1997) and has about 7000 nodes. In Figure 1 the node that has been just clicked by the user is *Artificial*. Only nodes that are close to the node in focus are visible.Whereas the areas of the tree around the point of interest are shown in great detail, there is also some sense of the overall structure of the tree. Because more information than actually requested by the click is displayed on the screen, the DOI trees may facilitate learning about the information hierarchy. Thus, tasks that make repeated use of the same hierarchy may benefit most from this kind of visualization.

Information Scent

One big component of the search heuristics used by people is information scent: when looking for a particular piece of information (target), a variety of semantic clues tell us how likely it is that the target be hidden under a particular label. These clues are referred to as ”information scent” (Pirolli & Card, 1999): when the information scent is high, people can find the target easily (e.g., Pirolli, Card, & Van Der Wege, 2000). Often it has been assumed that information scent reflects semantic similarity. Word similarity measures based on word co-occurrence statistics — such as Pointwise Mutual Information (PMI) (Turney, 2001) or Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997) — have been used suc-

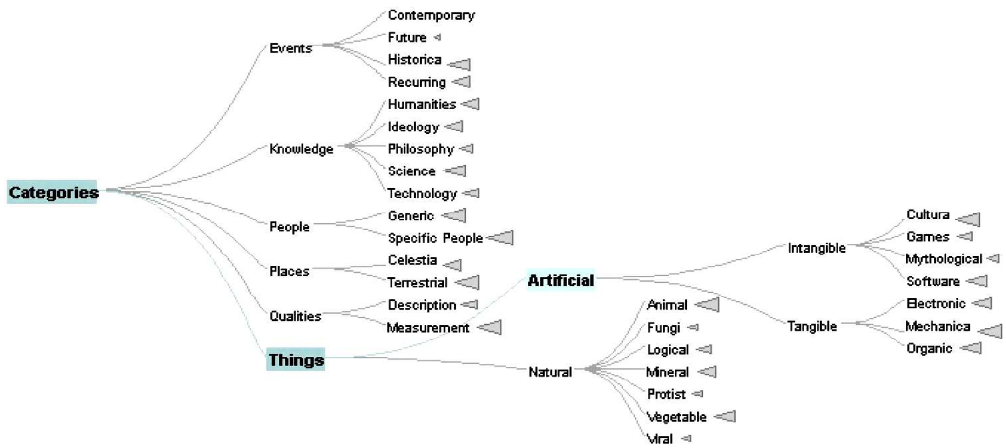

Figure 1: DOI Tree Visualization.

cessfully to predict information scent (Pirolli, 2005). However, such co-occurrence based metrics may not always be the best measures for information scent, especially when the information is organized taxonomically. For instance, when looking for the star *Alpha Centauri* in the structure in Figure 1, a node labeled *Wars* is not relevant, although *wars* and *star* have very high co-occurrence. In this case, a more appropriate measure for information organized taxonomically is degree of category membership: how much the target is a member of the class denoted by the label (Budiu, Pirolli, Fleetwood, & Heiser, 2006).

Previous work

The extensive literature on menu (or Web) navigation has focused on tree-like navigation spaces of limitted breadth and depth (e.g., Miller & Remington, 2004). Although some of these models have addressed the role of information scent (Miller & Remington, 2004; Fu & Pirolli, in press; Blackmon et al., 2005), they have not taken into account the role of the visual display and did not attempt to explain how information scent and visual organization interact to each other. Rather, they assumed a simple model of sequencing of attention to links on a display. Recent models of visual search of displays (e.g., Hornof, 2004; Tamborello & Byrne, 2005) address participant behavior on a single display, but do not deal with tasks that involve iterations of visually scanning a display, making a navigation choice, then repeating everything again over several more displays until the task is done. We fill the gap in the literature by presenting a model that (1) makes a sequence of navigation choices and solves a real search task; (2) takes into account both visual and semantic factors that affect navigation.

Experiment

Budiu et al. (2006) present an experiment in which people performed searches in DOI trees. Participants had to find 16 targets in a DOI tree visualization of the hierarchy in Figure 1. Half of the targets had high information scent (easy) and half had low scent (hard). In the high scent tasks, all the nodes on the path to the solution had highly salient labels. For the low scent tasks, some of the nodes on the solution path were less salient. For instance, the task to find the node *Banana* was high scent: the path to that node included highly salient items (*Categories; Things; Natural; Vegetable; Fruits; Tropical; Banana*); the task to find the node *Library of Congress* was low scent, because the path to that node (*Categories; People; Organizations; Governmental; United States; Legislative Branch; Library of Congress*) contained at least one less salient label (*People*). Budiu et al. (2006) report that people were faster in the high scent tasks (29.02s on average) than in the low scent tasks (63.27s on average). Participants tended to visit more nodes (85 nodes per task on average) and to wander farther away from the solution path (4.83 nodes away) in the low scent cases than in the high scent, where they visited only 47 nodes on average and departed from the solution path at a distance of 3.62 nodes on average.

The DOI-ACT Model

The DOI-ACT model is built within the ACT-R framework (Anderson & Lebiere, 1998). ACT-R is a cognitive architecture based on production systems; it has been used to model numerous experiments from various areas of psychology (problem solving, memory, language, etc.

The model has two main components that interact with each other: (1) a visual search component that parses the screen into visual groups and selects the most salient one to attend next, and (2) a semantic component that examines the nodes in the most salient visual group and decides on which one to click. Both these components use information scent to decide visual salience and semantic relevancy, respectively.

Measures of Information Scent

In our ACT-R model, information scent is ultimately mapped onto chunk activation: the more salient a node is to the current target, the higher its activation and the more easily it is retrieved from memory. The DOI-ACT model uses two measures of information scent (matched on two separate components of ACT-R activation): category-based scent and similarity-based scent. Category-based scent captures the degree to which a node is member of a category. The name of the category is called "hypernym" (e.g., *fruits* is a hypernym of *banana*). Although category scent plays the more central role in the model, occasionally similarity-based scent can be important, too. Indeed, nodes at higher levels in the ontological taxonomy (e.g., *Things*, *Fruits*) may be hypernyms of the target concept, rather than words semantically similar to that concept. However, the scent of the leaf nodes (e.g., *Pineapple*, *Banana*, *Mango*) with respect to a target such as *Banana* is better captured by similarity.

ACT-R activation has two components (Anderson & Lebiere, 1998): base-level activation, that corresponds to how often and how recently the chunk has been accessed, and a spreading activation, that indicates how much the chunk is associated with other chunks that are currently in the focus of attention (for instance, in the context). Category scent influences the base-level activation and similarity scent impacts spreading activation.

Category Scent. To estimate category scent, we use human category ratings for 1760 word pairs, collected via a web questionnaire. Seventy-one self-selected participants rated, on a scale from 1 to 5, how likely it is for a node (e.g., *banana*) to be member of a class (e.g., *fruits*). For a given search target, the category scent is used to set the ACT-R base-level activation of a special category chunk. For instance, to represent that *banana* is a *fruit*, we create a category chunk that links those two items and we assign to that chunk a base-level activation proportional to the average category rating for that pair.

Similarity Scent. To estimate similarity scent we used PMI (Turney, 2001) scores between the target (e.g., *banana*) and the nodes in the DOI tree. We computed PMI scores between all the nodes in the DOI tree and all the targets; however, for our model we only kept the top 100 PMI values (the others were so low that they could be considered zero). The PMI scores were used to set the ACT-R strengths of association S_{ij}: the higher the strength of association between *banana* and *mango*, the greater the activation spread received by *mango* from the target, and the higher the activation of *mango*. If *mango* had high activation when searching for *banana*, the model would infer that it was relevant for the current search goal.

Visual Search

There is a large literature in psychology that is devoted to visual search (see Wolfe, 1998 for a review). Many experiments studied how people search for a target when distractor items are present, and several theories of visual attention were proposed. DOI-ACT uses elements of CODE (Logan, 1996), a theory of visual attention that accounts for many phenomena in the literature. According to the CODE theory of

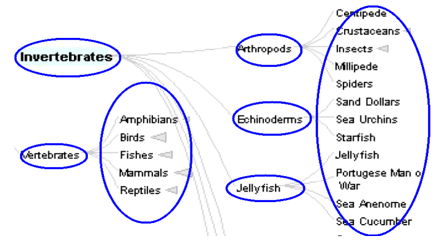

Figure 2: Different proximity-based groups as defined by our model.

visual search (Logan, 1996), people partition the display in several regions that are characterized by similar visual features, and then they select one of these regions. One such feature is proximity: items close together on the screen form groups. Figure 2 shows part of a DOI tree and the associated proximity-based groups.

Given a partitioning of the screen in visual groups, our model selects the most salient group and makes it the current group. Items in the current group are next examined in detail. There are many possible factors that can contribute to group salience (e.g., density, average information scent). To find out which of these factors really have an effect, we use a logit model. Logit models (and, more generally, discrete choice models) (Train, 2002) are a class of statistical techniques used primarily in economics to understand and predict how people select among several alternatives. We assume that visual groups are the alternatives (since people need to choose which one to attend next). We define a utility function (that corresponds to visual salience) for each alternative. The utility U (or salience) of a group g is a linear function of the variables $X_1 \ldots X_k$ considered:

$$U(g) = b_1 X_{1g} + b_2 X_{2g} + \ldots + b_k X_{kg} \tag{1}$$

where $b_1 \ldots b_k$ are some unknown coefficients.

According to the logit model, the probability of choosing one particular group depends on the salience of that group as in the following equation:

$$P(g) = \frac{e^{U(g)}}{\sum_{\text{all groups } k} e^{U(k)}} \tag{2}$$

We need to determine the coefficients $b_1 \ldots b_k$. Given that we know the choices made by the users (corresponding to the groups in which they clicked), we can determine these coefficients by using a maximum likelihood estimate (that is, we find the values for which the likelihood of the observed choices is maximum).

We considered several variables in the salience equation:

- horizontal distance between the center of the group and the current node[1]; that distance can be positive or negative, depending on whether the group is to the right or to the left of the current node;

[1]The current node is the node last clicked.

- Euclidean distance from the group center to the current node;
- number of descendants of the current node that are within the group;
- number of nodes in the group that were previously visited;
- number of nodes in the group that were previously clicked;
- density of the group (inverse of the number of nodes in the group);
- similarity-based scent (defined as the maximum similarity scent of all the nodes in the group);
- category-based scent (defined as the maximum category scent over all the nodes in the group);
- estimated category scent, defined as either an average of all category scents of previously visited nodes in the group, or, if no nodes were visited, the maximum category scent of all the parents for all the nodes in the group;
- estimated similarity scent: an average of all similarity scents for all the nodes in the group that were visited before;
- lag clicked: how long ago the model has already clicked an item in this group;
- lag visited: similar to lag clicked for visited nodes.

The estimated scent measures for a group reflect an expectation of the user that the group may be relevant, based on the scent of the nodes already observed and that are connected to it (e.g., parents of nodes in the group).

When we used the logit model, we obtained that most variables had close to zero coefficients in the utility equation. The only variables that had coefficients at least greater than 0.01 were horizontal distance ($b = 0.44$), Euclidean distance ($b = 0.1$), number of descendants ($b = 0.01$) in the visual group, category scent ($b = 0.034$).

However, when we tried a salience function that depended only on these variables we noted that (1) if we used both the Euclidean and the horizontal distance in the salience function, the distance would basically be the only variable that had an impact, (2) the model would tend to select the same groups over and over again, because no knowledge of what had been already visited was built in the salience function. Therefore, we updated the salience function to include the following factors:

- horizontal distance (D,)
- number of descendants (N) in the visual group,
- category scent (S),
- inhibition factor ($I = -100e^{1-c}$, where c is equal with the lag clicked) for the items that had been clicked recently.

Including the other variables with coefficients smaller than 0.01 did not affect the salience function[2]. Beside these variables, we added a normal noise component to the salience function. The normal noise injected randomness and non-determinism in the system.

[2] None of the variables had value ranges so big that even small coefficients could have led to them impacting the salience function.

Each time a new visual group must be selected by the model, this salience function is used. A parallel process calculates the salience of each visual group and the group with the highest salience is chosen. This salience function is fed into ACT-R's visual module and used each time a request for a visual group was made.

Note that, for groups placed to the left of the last node clicked (i.e., nodes upper in the tree), the distance is negative (and there are no descendants of the current node that belong to those groups), so category scent is really what makes the group salient. For nodes to the right of the last clicked node, the distance is positive and, together with the number of descendants, drives salience. Thus, people use category scent more in the upper levels of the tree, when they backtrack from the current node. They rely less on category scent as they advance deeper in the tree, by moving to the right of the current node. In fact, once they are on the right path, using distance as the main factor is a strategy that optimizes the time to the solution (the farthest away descendants of the current node would need to be clicked to get to the solution most quickly). It is when they made a mistake and needed to backtrack that users have to take scent into account more.[3]

Semantic Component of the Model

The model starts by selecting a group to examine. Once a group was chosen, the model scans the elements in the group one by one, top to bottom. In the original experiment, there was evidence that the participants preferred this sequential way of scanning. For each item processed, the model checks whether the node is a hypernym and marks it as such if so. When the model reaches the end of the group, it retrieves the best hypernym and clicks on it. For each group, the model starts with a high probability p of continuing and a small probability q of stopping. Each time an item of low relevance is encountered in the current group (i.e., an item with low category or similarity scent), the probability to continue p is decreased by a fixed value. When the probability of continuing becomes comparable with that of stopping, the model stops reading new items from that visual group.

Goodness of a hypernym translates into ACT-R activation. A category chunk such as "banana is a fruit" has an activation that corresponds to the degree to which *banana* is in the category *fruits*. Similarity-based scent is useful on leaf nodes of the tree, to decide whether a particular region in the space should be explored or abandoned (that is, if you are looking for *Banana* and encounter items such as *Jupiter* and *Mars*, you can safely guess that you are in the wrong region of the screen).

[3] Note also, that category scent is more available for the nodes to the left, because those nodes are "older" information that was semantically processed in the past, whereas the nodes to the right are newer nodes, whose semantic features may be less accessible on visual inspection.

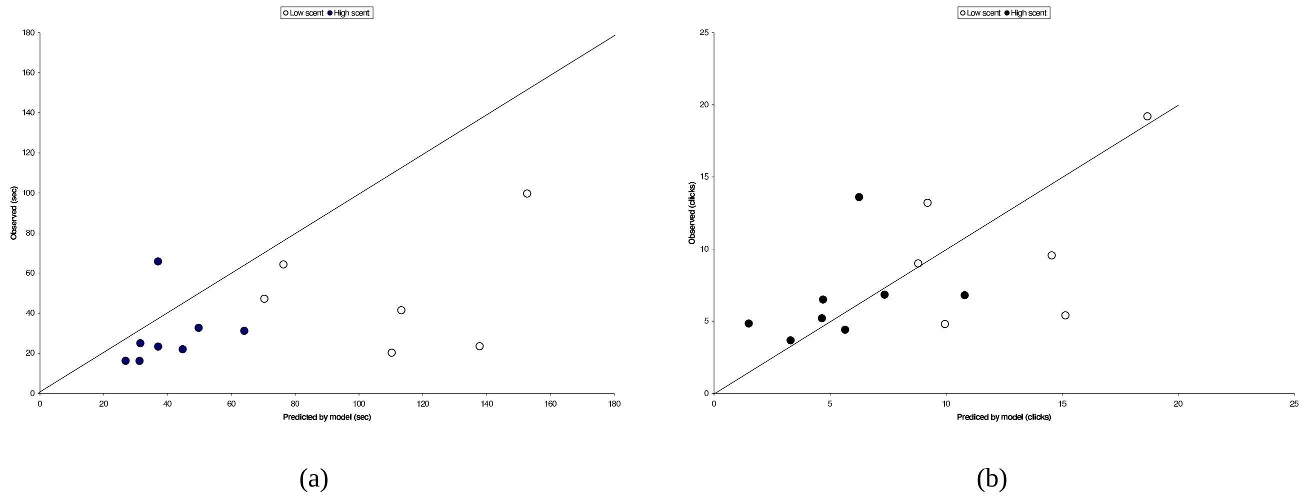

Figure 3: Model results (x axis) against data (y axis): (a) search times; (b) clicks.

Results and Discussion

The most important accomplishment of the DOI-ACT model is that it can successfully complete most of the tasks. The tasks in this model have an average solution path length of 8 nodes, with some of the tasks reaching 16. Even if the probability of making a mistake at one step were small, the overall probability of getting to the solution is low for such long solution paths. The fact that DOI-ACT is able to complete the tasks in reasonable time (less than 300 simulated ACT-R seconds) implies that we did capture a fair amount of the heuristics that people use to perform these tasks, and that the backtracking strategy that the model uses is fairly similar with that of humans.

DOI-ACT failed to complete two of the low scent tasks in less than 300s. However, these two tasks were the hardest for people as well (one participant failed to find the answer and all the others took very long time to complete these tasks, when compared to the other tasks).

The model fits the average behavior of the participants and the qualitative distinction between low and high scent: it took 119.15s (12.72 clicks) for low scent and 37.07s (5.52 clicks) for high scent, on average; people took 63.27s (10.19 clicks) for low scent and 29.02s (6.49 clicks) for high scent. DOI-ACT took longer than humans on low scent tasks, indicating that the backtracking heuristics used by people is not fully captured by the model.

Figure 3 shows how the model and the participants performed on the individual tasks. The model does a fairly good job on the high scent tasks, but captures less well some of the low scent tasks. There are at least two reasons for that. First, the low scent data has highly variable: there are 5 observations per task, and the search times for these observations vary widely between subjects. Second, although we used human ratings to estimate category scent, for some low scent tasks, this measure was not good enough, suggesting that people may pick up very quickly site-specific scent. For instance, the human participants in our rating study considered that *Celestial* is among the best hypernym for *god Ishtar*, much better than *People*. However, participants in the experiment never clicked on *Celestial* when searching for *god Isthar*. It is possible that they learned that *Celestial* in this taxonomy refers only to celestial objects, and not to deities. When we excluded the two low scent tasks where the category scent based on human ratings did not match our ontology, we obtained that the correlation between data and model was $R = 0.75$ for clicks and $R = 0.77$ for search times.

The results from the model indicate that, although the model is able to complete the task in reasonable time, it still takes longer than people on the low scent tasks, especially if these have a poor estimate of category scent. To improve the fit, there are several refinements that could be done to the search heuristics. One is to have a mechanism for learning site-specific category scent incorporated into DOI-ACT. Another avenue of future research is to further refine our visual search strategy with a theory of eye movements, to explain the actual fixations corresponding to the items processed by the model. In the future, we plan to use EMMA (Salvucci, 2001), a model of eye movements that works with ACT-R, to make that step.

Conclusions

We have presented DOI-ACT, an ACT-R computational model of how people navigate in a DOI tree. DOI-ACT has a visual component based on a discrete choice logit model, and a semantic component that uses information scent as ACT-R activation. It fits moderately well (in terms of clicks and search times) the data collected from human participants, suggesting that it can capture to a good degree the heuristics that people use in performing search in a large collection of information. However, the model tends to overestimate the time for search that people needed, indicating that some more work is needed to fully understand the strategies used by participants to backtrack from a wrong path.

One of the main contributions of this paper is in presenting a model that completes a complex, realistic search task on an interface that typically displays dozens of choices, in many different visual arrangements, dozens of times in a typical task. Moreover, the possible structure of behavior in this task is complex: an abstract representation of all the possible moves (in both attention space and navigation space), which could be made by either the users or the model, would reveal a state-space lattice of much greater complexity than that considered by previous navigation models (Miller & Remington, 2004; Blackmon et al., 2005; Fu & Pirolli, in press). Furthermore, in modeling any given DOI task, there are an extraordinary number of opportunities for the model to diverge from the user data when contrasted with previous models and approaches. In other words, there is a huge amount of items available on the screen and many bad choices to be made, and, if we want to do it in a cognitively plausible way, with cognitive operators (like those stipulated by ACT-R), not much time available. The combination of category scent and visual search strategy are really essential to capturing human data.

Another important contribution of this work refers to the discrete choice model of visual search in complex displays. We believe that this paradigm can be applied to study visual search in other complex interfaces, as well, and that it sets the basis of a microeconomics of the visual interface, placing us one step closer to the goal of building a science of complex visual interfaces.

Last, but not least, the DOI-ACT model brings forth a new understanding of information scent by distinguishing between category- and similarity-based scent. In terms of ACT-R modeling, in the past, information scent was mapped on production utility (Pirolli et al., 2000); we diverted from that tradition and incorporated scent more naturally in the declarative memory module, as chunk activation.

In conclusion, we believe we are in a better position to give some guidance to interface designers. One principle that seems to come out of this work is that the effect of information scent cannot be overestimated. Our model did a good job of capturing high scent items and had a harder time with low scent items, and so did people. Well chosen labels, that truly reflect what is hidden behind them and do not side-track users, are essential for rapid search. Moreover, our research shows that labels with strong categorical scent are better placed close to the root of the hierarchy.

Another principle, that comes out directly from Equation 2, is that the more visual groups on the screen, the harder they have to work to become salient: too many groups on a display will just dissipate users' attention, irrespective of how clever the display is.

Acknowledgments

Portions of this research have been funded by an Advanced Research and Development Activity, Novel Intelligence from Massive Data Program, Contract No. MDA904-03-C-0404. We thank Stuart Card for comments on an earlier version of the paper.

References

Anderson, J., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, New Jersey: Lawrence Erlbaum.

Blackmon, M., Kitajima, M., & Polson, P. (2005). Web interactions: Tool for accurately predicting website navigation problems, non-problems, problem severity, and effectiveness of repairs. In *CHI'05*.

Budiu, R., Pirolli, P., Fleetwood, M., & Heiser, J. (2006). Navigation in degree-of-interest trees. In *AVI'06*.

Card, S., Mackinlay, J., & Schneiderman, B. (1999). *Information visualization*. Morgan Kaufmann.

Card, S., & Nation, D. (2002). Degree-of-interest trees: Component of an attention-reactive user interface. In *AVI'02*.

Fu, W.-T., & Pirolli, P. (in press). SNIF-ACT:A cognitive model of user navigation on the world wide web. *Human-Computer Interaction*.

Hornof, A. (2004). Cognitive strategies for the visual search of hierarchical computer displays. *Human-Computer Interaction*, *19*, 183-223.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*, 211-240.

Logan, G. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, *103*, 603-649.

Miller, C. S., & Remington, R. W. (2004). Modeling information navigation: Implications for information architecture. *Human-Computer Interaction*, *19*(3), 225-271.

Mullet, K., Fry, C., & Schiano, D. (1997). On your marks, get set, browse! In *CHI '97*.

Pew internet survey. (2006, December). http://www.pewinternet.org/. (Downloaded 1/24/07.)

Pirolli, P. (2005). Rational analyses of information foraging on the web. *Cognitive science*, *29*(3), 343-373.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*.

Pirolli, P., Card, S., & Van Der Wege, M. (2000). The effect of information scent on searching information visualizations of large tree structures. In *AVI'00*.

Salvucci, D. (2001). Integrated model of eye movements and visual encoding. *Cognitive Systems Research*, *1*, 201-220.

Tamborello, F., & Byrne, M. (2005). Information search: the intersection of visual and semantic space. In *CHI'05*.

Train, K. (2002). *Discrete choice models with simulation*. Cambridge University Press.

Turney, P. (2001). Mining the web for synonyms: PMI-IR versus LSA on TOEFL. In *ECML-2001*.

Wolfe, J. (1998). Attention. In H. Pashler (Ed.), (chap. Visual Search). University College London Press.