

Modeling Web Navigation: Methods and Challenges

Craig S. Miller

DePaul University
cmiller@cs.depaul.edu

Abstract. A computational cognitive model of Web navigation is a working computer system that simulates human users searching for items in a Web site. A fully working model must automate aspects of human perception, decision making and physical control. To successfully predict human behavior, these automated processes must be consistent with the cognitive and physical limitations of human users. Predicted behavior might include which links users select, when they select them and when they backtrack to previous pages. In this chapter, the necessary capabilities of a working model are described in detail. These include processes that simulate users scanning a page of links, assessing each link, selecting a link and deciding when to backtrack. Accurately modeling link assessment for a variety of users is critical for successful predictions and is perhaps the greatest challenge in creating a useful model. Several approaches to link assessment are presented. The implementation details of one model are described, which are then evaluated by correlating the model's timing predictions to results from user studies.

1 Introduction

Cognitive models account for human behavior from an information-processing perspective. As a functioning computer program, these models simulate aspects of human perception, cognition and decision making as they accomplish some task. In this way, it offers an account of how humans perform the task. In this case, the task is Web navigation and the goal is to simulate human users navigating a Web site in order to find the items they are seeking.

Web navigation is perhaps the most common strategy for finding an item in a Web site (Katz & Byrne, 2003). Sometimes called “browsing,” it involves identifying relevant links on a Web site and selecting those that will likely lead to the sought-after item. Usually several iterations of page scanning and link selection are required before the targeted item is found. Often some backtracking is needed for cases when misleading links are selected.

A contrasting strategy is the use of a site's keyword search facility. Sometimes simply called “search,” this method requires the user to specify some query terms that hopefully identify the user's content goal. The site's search facility then returns a page of links relevant to the specified terms. Keyword search requires the user to *recall* relevant terms whereas the principal cognitive skill for Web

navigation requires the user to *recognize* relevant terms. While both strategies have their uses, some empirical studies suggest that users more frequently employ Web navigation and, when they do, are more likely to find the targeted item (Katz & Byrne, 2003; Campagnoni & Erlich, 1989).

Both Web navigation and keyword search have an assumption that the user has some kind of content goal that may be ultimately fulfilled on a content page. This goal may be well defined in the form of a specific object or it may refer to a general category of items. This assumption may not be true for some styles of Web interaction. For example, users may choose to interact with a Web site to merely review its contents.

While Web navigation does not account for all the ways in which users may interact with a Web site, it is perhaps the predominate activity. Web designers develop Web sites with the aim of supporting effective Web navigation. With this in mind, a model of Web navigation provides a substantial account for how users interact with Web pages and is of considerable interest to those who design them.

A cognitive model of Web navigation is useful in a variety of ways. Assuming the model reasonably approximates human usage, it predicts human behavior. Its predictions indicate which links users will select and when users backtrack to previous pages. By assigning time costs to its actions, the model can predict the time required to find a targeted item. With little cost, the model can provide predictions for a range of parameters on a variety of structures. The working model can run in place of user studies. Unlike user studies, the model also offers an explanation of human behavior. This insight allows a designer to understand the impact a change in design might have or how a result might generalize to other structures.

In this chapter, I review the Web navigation task and discuss approaches to model it. I start by describing the cognitive activities that support Web navigation and discuss several approaches to modeling each. Then I review one particular model called MESA, including what it models, some evaluation results, and how it has been useful. I present some challenges that still need to be addressed. Finally, I discuss some of MESA's implications on intelligent systems that try to infer the intent of users.

2 Performing Web Navigation

A complete cognitive model of Web navigation must account for the following activities that support the task:

- Visually scanning links on a page
- Assessing links with respect to the user's navigation goal.
- Selecting links.
- Assessing when to return to a previous page to attempt an alternate path.

For now, it will be convenient to analyze these activities separately, but later we will consider examples where these activities strongly interact and mutually determine what strategies are used.

2.1 Visually Scanning a Web Page

When users encounter a navigation page, they typically identify the links on the page and sequentially attend to them. In the simplest case, the page provides a serial list of links that imply a logical order in which they should be evaluated. For example, the links could be arranged vertically from top to bottom and thus imply the order in which many users would scan them.

While many Web pages use a simple serial layout, perhaps most pages present an arrangement of links whose spatial placement by itself does not imply a scanning order. Visual attributes such as motion (e.g. blinking elements), large fonts or bright colors generally attract a user's attention. Faraday (2000) has incorporated these visual attributes into a working model that indicates a plausible order in which users would scan a page. The model also follows a left-to-right, top-to-bottom scan after the starting point has been determined.

Often a Web page groups links that are related to each other and labels the grouping with a higher level category. In these cases, users usually choose to scan the group labels before they consider the link selections within each group (Hornof & Halverson, 2003).

Users' experiences with Web sites may also influence the order in which they scan a page. For example, many users have learned that the content in banner advertisements do not contain content that interest them. Consequently many users skip their links (Benway, 1998) even if they cannot entirely ignore them (Burke, Gorman, Nilsen, & Hornof, 2004).

Also, many Web sites have adopted a consistent scheme for placing links on pages. Top level categories may be placed horizontally at the top of the page and secondary categories may be displayed vertically at the left side of the page. Users who learn this scheme may only scan the links pertinent to their navigation goals.

2.2 Link Assessment

As users attend to each link while scanning a page, they assess the link label with respect to their navigation goal. They gauge how likely the link will lead to the target. Various terms have been used for this subjective measure. They include residue (Furnas, 1997), relevance (Young, 1998) and information scent (Pirolli & Card, 1999).

Sometimes the assessment is trivial. A user may see the exact text or image that precisely matches the navigation goal. In these cases, the assessment can be made based on the superficial properties of the label. For example, if the user is looking for bicycles, the string "Bicycles" or an image of a prototypical bicycle immediately indicates that selecting this link will lead to these items.

Other times the user needs to assess the link label as a category and evaluate the extent to which the user's navigation goal belongs to the category. For example a search for bicycles may involve assessing a link labeled with "Sporting Goods." A useful proxy for category membership is semantic similarity. Analysis tools such as Latent Semantic Analysis (LSA) produce a similarity metric for

a pair of phrases (Landauer & Dumais, 1997). These pairs may correspond to the navigation goal (e.g. bicycles) and the label (e.g. “Sporting Goods”) and the resulting similarity metric can indicate the label’s relevance. LSA has been used to evaluate label quality in Web applications (Blackmon, Kitajima, & Polson, 2003). Later in this chapter, I will further discuss various approaches for estimating link relevance.

2.3 Link Selection Strategies

Assessing relevance of a link label may not in itself determine whether the link will be selected. A link selection strategy may depend on a threshold. If the relevance is above an established threshold, the link is selected. Otherwise, scanning proceeds to the next link for assessment. The threshold may be lowered for a secondary pass if the first pass failed to identify successful links.

An alternate selection strategy is link comparison. A user may first assess several links and then select the link with the highest relevance. Cognitively, the comparison strategy requires more resources than the threshold strategy since the user must remember the highest link value and where the link is located (Miller & Remington, 2004). Determining which strategy is more efficient depends on the quality and distribution of the links labels. For a well designed page that has one relevant link and no misleading links, the threshold strategy is more efficient since the relevant link is selected as soon as its label is assessed.

So far, we have assumed that these three abilities are modular activities. However, they may interact in practice. For example, if the user already knows the visual features for identifying the label, the user may use a preattentive search strategy. For some visual features such as color, the desired target draws the user’s attention without requiring a serial scan (Triesman & Souther, 1985). In one study, users were able to identify and select the color-coded link without scanning and assessing the other links on the page (Ehret, 2002). In another study, there is evidence that users can evaluate multiple targets at once if they know the actual text of their target (Hornof & Halverson, 2003).

2.4 Backtracking Strategies

Sometimes users select links that do not lead to the target. With this possibility in mind, they must continually reassess if they are pursuing a path that will lead them to their goal. For one strategy, they may decide to return to the previous page when they no longer see links whose relevance exceeds their selection criterion. Alternatively, they may choose to lower their criterion momentarily to remove any doubt that they have made the wrong selection. In previous work, Roger Remington and I provide accounts of this second strategy (Miller & Remington, 2002). We call this the *opportunistic strategy* since users can explore less likely options on the current page while the opportunity presents itself. If none of these options succeed, they can confidently rule out this path and return to the previous page.

3 Modeling Approaches

There are no comprehensive cognitive models of Web navigation. Some models have focused on how users scan menus (Byrne, 2001; Hornof & Halverson, 2003) and Web pages (Faraday, 2000). Other models account for how users navigate through a sequence of pages (Lynch, Palmiter, & Tilt, 1999; Chi, Rosien, Supattanasiri, Williams, Royer, Chow, Robles, Dalal, Chen, & Cousins, 2003). While some of the models operate on actual Web sites (Lynch et al., 1999; Chi et al., 2003), none consider the visual attributes in a page for constraining the order in which links are evaluated and selected.

Usually the construction of a cognitive model starts with idealized assumptions. For example, the Max Model (Lynch, Palmiter, & Tilt, 1999) assumes that the user always selects the links that lead to the target. It predicts navigation times by summing typical costs of the user actions needed to reach the target. Card, Moran and Newell (1983) show that this approach can be effective for predicting task completion times of practiced users, but it is not clear how well the predictions of the Max Model correspond to actual users navigating a Web site (Pirolli & Card, 2000).

Other approaches use bounded rationality as a guiding principle for constructing a working model. Here the assumption is that people generally act in ways that will efficiently achieve their goals, at least within the bounds of their knowledge, cognitive resources and physical abilities. Bounded rationality and its variants have a long history for their successful application in predictive models (Simon, 1981; Anderson, 1990). Rational analysis has led to predictive models of information foraging (Pirolli & Card, 1999). For Web navigation, the SNIF-ACT model is constructed using ACT-R, a cognitive architecture motivated by rational analysis (Pirolli & Fu, 2003). MESA is a cognitive model of Web navigation I developed with Roger Remington (Miller & Remington, 2004). In the next section, I describe our approach to using principles of bounded rationality and abstraction in developing the model. A more detailed description with extensive traces appears in Miller and Remington (2004).

4 The MESA Model

In developing MESA, we followed three principles:

- The limited capacity principle
- The simplicity principle
- The rationality principle

Combining the limited capacity principle and the rationality principle is our approach to bounded rationality. We construct the model using navigation strategies that minimize cognitive resources. For example, as we have noted, the threshold strategy for selecting links minimizes memory requirements (limited capacity principle) while also being effective for most scenarios (rationality principle).

We are also interested in a model that is reasonably easy to understand, implement and replicate. With the simplicity principle, we seek a model that approximates human behavior to produce useful predictions while avoiding complexities that provide only marginal benefits. For example, MESA assumes a fixed time cost for assessing the relevance of each link. While actual time costs certainly vary as a function of label length, label relevance and the user’s reading ability, it is not clear if an account of these variations would produce substantially better predictions since the variations may average out in actual usage.

4.1 Modeling the Web Site and Web Browser

At this time, MESA does not interact with actual Web sites. Instead, its simulations run on abstract representations. While these representations do not specify the visual attributes of the site’s pages, they indicate site’s abstract structure, often called the information architecture.

Figure 1 shows a simple example of an abstract site structure. The rectangle at the top represents a starting page with four links. Each of those links leads to additional pages, each with two links. The shaded rectangle represents the page that has the user’s target or navigation goal. Each link is represented with a number from 0 to 1. Based on the label for the link, this number represents the user’s subjective assessment of how likely selecting the link will lead to the target. This assessment is made independent of the other links on the page. In this way, a page may have any number of links with relevance assessments close to one.

The numbers correspond to one user’s assessment. A different user may evaluate the link labels in a different way. Thus, this representation models a site and a user’s interpretation of its link labels with respect to the user’s navigation

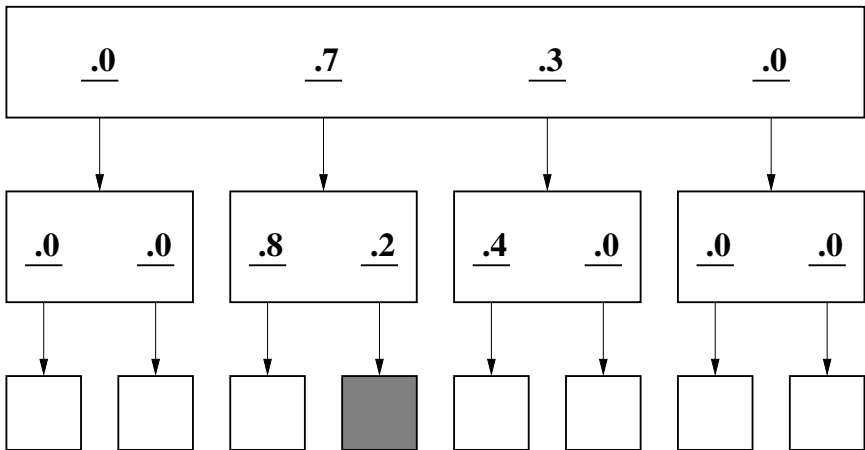


Fig. 1. Abstract representation of a Web site

goal. In the next section, I will review some methods for obtaining relevance assessments based on real link labels.

These structures can represent different cases. For a well designed Web site, the most relevant links lead to the user's goal. This is the case for the top level page for the site in Figure 1. However, the second page (with links marked .8 and .2) has a misleading link. That is, a highly relevant link (.8) does not lead to the target. For this representation, the user must select the less relevant link (.2) in order to find the target. Figure 1 thus represents well designed links at the top level and a misleading link at the second level. By manipulating the values that correspond to relevance assessments, this representation scheme can model well designed sites, where the most relevant links lead to the navigation goal, and poorly designed sites, which consist of many misleading links.

We simulate the most common operations of a Web browser:

- Selecting a link
- Pressing the back button
- Highlighting which links have already been selected

In one study of Web usage, selecting a link and pressing the back button comprise of more than 80% of Web navigation actions for viewing a page.

4.2 MESA Strategies

Applying our principles of Web design, we have implemented the threshold strategy for selecting links and the opportunistic strategy for temporarily delaying when to retreat to a previous page. For this strategy, the threshold may be lowered for a secondary pass in order to explore marginally relevant links before it returns to the previous level.

At this time, MESA makes no commitment on the order in which links are scanned. Eventually we would like to incorporate visual rules that determine this order. For now, we either work with simple designs that imply an order (e.g. a list running from top to bottom) or we randomize the order in our simulations.

The flowchart in Figure 2 summarizes the threshold and opportunistic strategies. Starting with a new page, it scans links in a serial order. If it assess a link whose relevance exceeds the threshold, that link is selected. Otherwise, it continues scanning and assessing links. When it reaches the end of the page, it rescans the page a second time with a lower threshold unless one of the following is true:

- The threshold has already been lowered.
- The model can determine that it did not encounter any marginally relevant links on the first scan.

To recall the presence of a marginally relevant link, the model keeps one memory bit that indicates whether such a link was encountered. To model the short-term memory of a human (limited capacity principle), it loses this memory when it selects a link to scan a new page. In these cases, the model may perform

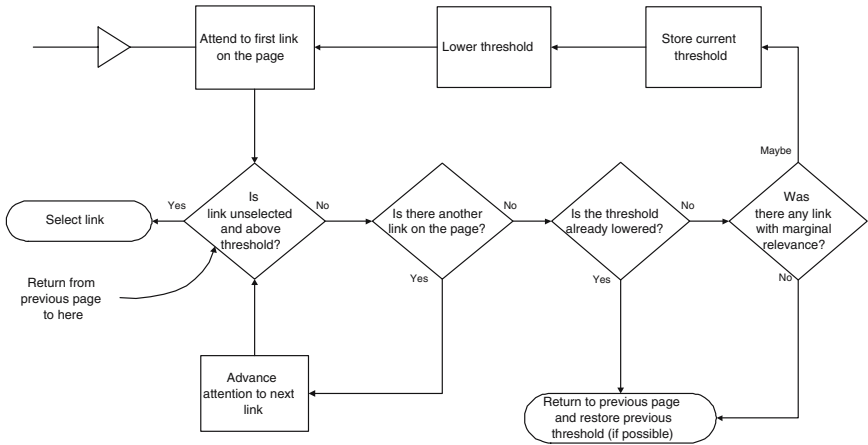


Fig. 2. Flowchart summarizing MESA’s strategies

a second scan when it returns even if it had not encountered any marginally relevant links on this page.

The model also maintains a limited memory for the threshold at a previous level. If the memory capacity has not been exceeded, the model can restore its threshold to the pre-existing value. We have explored the effects of various memory limits elsewhere (Miller & Remington, 2004).

To illustrate how MESA navigates an example structure, let us consider Figure 1. For this example, we will use .5 as the initial selection threshold. Starting with the first link at the top level (.0), its value is not close to the threshold. Link scanning proceeds to the second link (.7). This value is above the threshold and the model selects it. The next page appears. The first link on this page (.8) is above the threshold. The model selects it but finds that this link does not lead to the target. Returning to the second level, the model scans the next link (.2) but does not select it because it is below the threshold. However, before the model returns to the top level, the opportunistic strategy temporarily reduces the threshold to .1. Rescanning the page, it then selects the second link and finds the target. If the model had not found the target, it would have returned to the top level and continue scanning using the original threshold. At this level, it may reduce the threshold after the initial scan if it still does not find the target.

As the model simulates Web navigation, it employs three basic operations:

- Assess the relevance of a link label
- Select a link
- Return to the previous page (i.e. press the “Back” button)

By assigning plausible time constants to these operations, the model can predict the amount of time needed to find targets in a Web structure. In one of our studies, we obtained good absolute fits with 500 milliseconds for assessing a

link's relevance, 2.5 seconds for selecting a link and 1.5 seconds for returning to the previous page.

Not every user employs the threshold and opportunistic strategies like our model. For example, Howes, Payne and Richardson (2002) provide evidence that some users consider the relevance of competing links at the top level for deciding whether to pursue less relevant links at a secondary level. However, it is less clear if a substantial number of users employ this strategy under most circumstances. In the case of the Howes *et al.* study, menu structures consisted of only 2 links per page. Moreover, it appears that many participants might have used strategies consistent with MESA. With this mind, our approach is to continue with the simpler modeling strategies as long as they produce reasonably accurate timing predictions.

Despite the simplicity of MESA's strategies, the model accounts for a broad range of user behavior. If each visited page contains one link with high relevance and the remaining links with low relevance, the model quickly navigates deep into the Web site's structure. Alternatively, if the top level page has several links with high relevance each leading to pages with low relevance links, the model exhibits shallow exploration. Intermediate exploration behaviors are created by modeling various distributions of link relevance across the pages. Thus, the simple threshold and opportunistic strategies yield a range of behaviors as determined by link relevance. Our working hypothesis is that these strategies provide a useful approximation for understanding most users as they navigate Web sites. We may ultimately consider more sophisticated strategies but only if they achieve substantially better correlations to human behavior.

4.3 Assessing Link Relevance

For its simulations, MESA requires relevance assessments for the links in the structure. One method for setting relevance values involves starting with a Web site whose links are ideally labeled. That is, the links that lead to the target are valued at 1.0 and the links that do not lead to the target are valued at 0.0. This structure represents a perfectly designed Web site for a user and a navigation goal. We can then create less ideal sites by adding random noise to the link values. Adding more noise increases the probability that the model selects misleading links, that is, links that do not lead to the target. We have used this approach to show how the model replicates results in a user study that compared the effectiveness of three different structures (Larson & Czerwinski, 1998). Here effectiveness is defined as the average time required to find targets. With sufficient noise, the model's simulated times ranked the effectiveness of the structures in the same order as that of the user study.

A second approach is to employ human raters. For the results in Miller and Remington (2004), we and a third judge rated text labels with respect to a target. Each label-target pair received one of three discrete ratings: probably lead to the target, possibly but unlikely lead to the target, and highly unlikely to lead to the target. By assigning values to these categories (respectively 1.0, 0.5 and 0.0), we

obtain numerical averages and variations. More recently, I have collected ratings from larger samples. I will review these results in the next section.

Assigning random noise to ideal relevance values is a low-cost method, but it assumes a general level of quality throughout the site. The use of human raters provides a better assessment but requires costly procedures for collecting the ratings. We are considering automatic methods for assessing label relevance, which I further explore later in this chapter.

5 Evaluation

We have assessed MESA’s validity by comparing its predictions to results from human user studies (Miller & Remington, 2004). At the abstract structural level, we have reported how MESA produces results that are consistent with those in several menu selection studies (Miller, 1981; Snowberry, Parkinson, & Sisson, 1999) as well as the Web study by Larson and Czerwinski (1998). We also report our own user study where we asked human participants to find 8 department store items in structures consisting of nearly 500 items. We used our judged ratings (3 raters) to create relevance values for the MESA simulations. To account for variation among the site’s human users, we ran the simulations on structures that had relevance values spread along a normal distribution as defined by the average and standard deviation of the judged assessments. In this way, the simulations always used the same relevance values for when the judges agreed on the value but used a range of relevance values for when the judges disagreed. The simulated times were then averaged across 1500 runs (100 for each human participant) producing a total of 24 predicted times (8 targets on three different

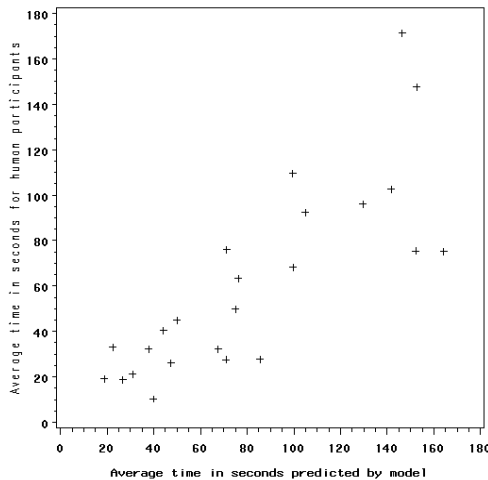


Fig. 3. MESA’s predictions compared to actual navigation times

structures). These simulations produced timing predictions that had a Pearson correlation of 0.79 to the navigation times.

While there is some precedent for using expert judgments to evaluate user interfaces (e.g. see Nielsen and Mack, 1994, for a collection of methods involving expert assessments), it is not clear how well these judgments match the behavior of real users. With the goal of obtaining more accurate relevance assessments, I asked 17 human participants to rate the labels for each of the 8 targets. These participants were recruited using the same method as the user study. Even though there was some disagreement between these ratings and the judged ratings (the Pearson correlation is .74), the simulated predictions from these 17 participants only produced a marginally greater correlation (.81) to the navigation times from the user study. Figure 3 shows how the predicted times of the model match those produced by human users. The x-axis shows the times predicted by MESA and the y-axis shows the average time for the participants. To produce the simulated times, the time costs for evaluation, selection and return were respectively 500 milliseconds, 2.5 seconds and 1.5 seconds.

At this time, it is not clear if better predictions will come with more accurate assessments of label relevance or by revising the model so that its strategies better match those of human users. A future goal is to obtain better relevance assessments by having the same human users provide the assessments and the navigation times.

6 Future Directions for Assessing Relevance

Collecting assessments of label relevance from a large number of users is reasonable and necessary for validating the model. However, the use of a large number of human raters is too costly as a routine method for evaluating the quality of a Web structure. Asking human users to rate labels requires about as much time as asking them to perform the actual navigation tasks. Using a small number of expert judges is more feasible but runs a greater risk of being less reliable. Even with a reduced number of human raters, the expense of an exhaustive assessment is considerable. For example, a site with 1000 targets and 100 category labels requires 100,000 assessments.

While it may be possible to effectively use a more manageable representative sample of human-rated assessments, a simulation performing a comprehensive evaluation would need to use some kind of automatic method. At this time, there does not appear to be any method that is currently able to replace human assessments of label relevance. Here I review some approaches, their current shortcomings and possible directions for improvement.

6.1 Co-occurrence of Label and Target

If one has access to documents that contain the site's labels and targets, one can use the frequency of how often the label and target co-occurs as the basis of a similarity measure. A useful measure results by normalizing the frequency with

respect to how often the label and the target appear independently. Pirolli and Card (1999) have successfully used co-occurrence to predict user behavior in an information foraging task. More recently, Pirolli and Fu (2003) used the Web for their model of Web navigation.

Successfully using co-occurrence requires that the label and the target frequently appear in the body of documents. Web search engines provide access to a large number of Web pages increasing the likelihood that full label and target names appear with a sufficient frequency. However, some specialty items may still not appear with enough frequency. For example, the site for our user study has the item “Tripod grill” and its relevance to the category “BBQ Tools and Gadgets” needs to be assessed. Unfortunately a Web search engine did not identify many pages with the full names: 94 pages for “Tripod grill”, 3 pages for “BBQ Tools and Gadgets” and 1 page for them combined. These numbers are too small to reliably assess relevance. Of course, stemming the words and breaking up the phrases would increase the number of matches, but this would then lose contextual information.

6.2 Latent Semantic Analysis

Latent Semantic Analysis (LSA) provides a similarity measure that could be used for relevance. Like counting co-occurrences, it depends on a body of documents. However, its measure is not fully determined by how often the label and target appear together. Also word order is not considered. Instead it treats the component words as vectors and vector similarity depends not just on co-occurrence but also how often related words appear together.

Without word order as a consideration, LSA requires extended text to provide the context. Blackmon, Kitajima and Polson (2003) used LSA for diagnosing usability problems in Web sites by describing the navigation goal with a lengthy segment of text (100–200 words). However, the shorter names of labels and targets in the domain of our user study do not appear to be sufficiently rich for successfully applying LSA. When examining the text labels that led to the navigation goals of our user study, the correlation between the relevance ratings produced by LSA and the judged ratings was only .28 and not significant ($p = .18$). These LSA similarity ratings were calculated using the LSA Web site (<http://lsa.colorado.edu>, accessed February 3, 2003). It is probable that better correlations would come from an analysis based on a larger body of documents containing text more relevant to our domain.

6.3 Wordnet

One shortcoming of using a similarity metric is that it does not distinguish between category membership and other similarity relationships. For example, most users would not select a link labeled “Tricycles” in order to find bicycles. Yet, a simple similarity measure would indicate a high level of similarity between this label and this navigation goal.

Wordnet is an online database of words and phrases that provides relational connections between the entries (Miller, 1995). Among the connections is the

hyponymy connection denoting category membership. This relation could be useful in distinguishing between general similarity and category membership. While there exist Wordnet-based measures that could serve as relevance measures (see Pedersen, Patwardhan, and Michelizzi, 2004, for a summary), measures of category membership that could extend to assessing relevance have yet to be developed. Moreover, additional work is needed to extend the database to labels that are not explicitly noted in the database.

6.4 Modeling Variation Among Users

For any of these approaches, it is not enough for them to return measures of relevance that would be useful for information retrieval. If our goal is to model users, a valid method must be able to model a range of users. With human raters, we can simulate user variation by creating multiple sets of assessments whose levels vary as a function of the variation in the human ratings. For automated similarity measures, we need approaches for simulating this variation. For example, we may find that the least common terms in corpora produce the greatest variation of assessments among humans. If so, frequency of occurrence may serve as a good source for simulating assessment variation.

7 Implications for Intelligence Systems That Infer User Intent

Our results indicate that MESA's opportunistic strategy is a useful describing human navigation patterns. This has implications for an intelligent system that tries to infer a user's intent based on the user's link selections. This includes systems that analyze Web server logs with the goal of determining what the user was looking for. It also includes recommendation systems.

Figure 4 shows an example structure where a user's link selections may mislead a system that is attempting to infer the user's intent. Scanning from left to right, the user would select the second link in Page A (valued at .6). Page C is then scanned. On the first pass, no links are selected, but MESA predicts that the user will perform a second pass at a lower threshold that would cause all of these links to be selected. The user returns to Page A only after learning that all of these links do not lead to the selected target. If the user's memory allows it, the threshold is restored to its original value (e.g. .5) when returning to Page A. Otherwise, the third link on Page A (valued at .2) might also be selected as well as two links on Page D. It is only after these choices are exhausted would the user select the highly relevant links that lead to the target (assuming that the user has not given up).

This navigation of the site in Figure 4 illustrates how users may select many links whose labels are only marginally relevant to their navigation goals. In this way, most of the selected links are not good indicators of user intent. If an intelligent system were to consider the site structure and the sequence of link selections, it might do well to disregard (or at least minimize the influence of)

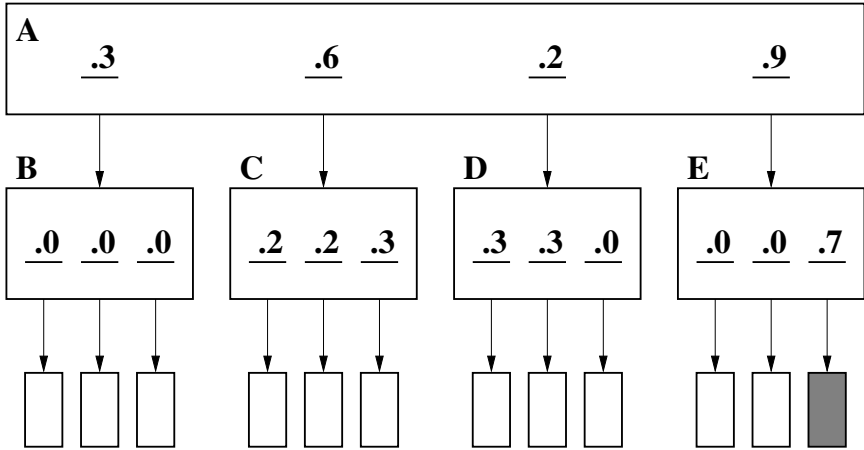


Fig. 4. Example structure that may mislead intelligent systems

link selections made on lower pages if the user eventually returns to a higher level to try other links.

8 Closing Comments

In the long term, we would like to see systems like MESA simulate users on actual Web sites with the goal of providing useful feedback on the accessibility of the site’s content. In the place of costly user studies, the model would simulate users and indicate where real users would encounter difficulties. Of course, the construction of this kind of model would need to overcome some challenging obstacles, some of which I have reviewed here.

Still, even in its current form, models like MESA have been useful. Already we have applied MESA to resolve issues addressing the optimal number of links per page. We also envision using MESA early in the design process to compare the effectiveness of different structures. At this point in the development process, the abstract representations are appropriate and relevance values can be determined using a combination of methods. Finally, models like MESA give us an understanding of human behavior. This insight has a range of uses including a better understanding of what may be inferred when a user selects a link.

Acknowledgements

The author thanks Sven Fuchs and Sarabjot Anand for their helpful comments on earlier versions of this chapter. Funding for this work was provided by the Human-Automation Integration Research portion of the Airspace Operation Systems Project of NASA’s Airspace Systems Program.

References

- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Benway, J. P. (1998). Banner blindness: The irony of attention grabbing on the world wide web. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 463–467).
- Blackmon, M. H., Kitajima, M., & Polson, P. G. (2003). Repairing usability problems identified by the cognitive walkthrough for the web. *Proceedings of the conference on Human factors in computing systems* (pp. 497–504). ACM Press.
- Burke, M., Gorman, N., Nilsen, E., & Hornof, A. (2004). Banner ads hinder visual search and are forgotten. *Extended Abstracts of ACM CHI 2004: Conference on Human Factors in Computing Systems*. ACM Press.
- Byrne, M. D. (2001). Act-r/pm and menu selection: Applying a cognitive architecture to hci. *International Journal of Human-Computer Studies*, 55, 41–84.
- Byrne, M. D., John, B. E., Wehrle, N. S., & Crow, D. C. (1999). The tangled web we wove: A taskonomy of www use. *Proceedings of CHI'99 Human Factors in Computing Systems* (pp. 544–551). New York: ACM Press.
- Campagnoni, F. R., & Erlich, K. (1989). Information retrieval using a hypertext-based help system. *Proceedings of the 12th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 212–220). ACM Press.
- Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Chi, E. H., Rosien, A., Supattanasiri, G., Williams, A., Royer, C., Chow, C., Robles, E., Dalal, B., Chen, J., & Cousins, S. (2003). The bloodhound project: automating discovery of web usability issues using the infoscentp simulator. *Proceedings of the conference on Human factors in computing systems* (pp. 505–512). ACM Press.
- Ehret, B. D. (2002). Learning where to look: Location learning in graphical user interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 211–218). ACM Press.
- Faraday, P. (2000). Visually critiquing web pages. *The 6th Conference on Human Factors & the Web*. Retrieved from <http://www.tri.sbc.com/hfweb/faraday/FARADAY.HTM> on April 19, 2004.
- Furnas, G. W. (1997). Effective view navigation. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 367–374). ACM Press.
- Hornof, A. J., & Halverson, T. (2003). Cognitive strategies and eye movements for searching hierarchical computer displays. *Proceedings of the conference on Human factors in computing systems* (pp. 249–256). ACM Press.
- Howes, A., Payne, S. J., & Richardson, J. (2002). An instance-based model of the effect of previous choices on the control of interactive search. *Proceedings of the 24th Annual Meeting of the Cognitive Science Society* (pp. 476–481). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Katz, M. A., & Byrne, M. D. (2003). Effects of scent and breadth on use of site-specific search on e-commerce web sites. *ACM Trans. Comput.-Hum. Interact.*, *10*(3), 198–220.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.
- Larson, K., & Czerwinski, M. (1998). Web page design: Implications of memory, structure and scent for information retrieval. *CHI'98: Human Factors in Computing Systems* (pp. 25–32). New York: ACM Press.
- Lynch, G., Palmiter, S., & Tilt, C. (1999). The Max model: A standard web site user model. *Proceedings of the 5th Annual Human Factors and the Web Conference*. Retrieved from <http://zing.ncsl.nist.gov/hfweb/proceedings/lynch/index.html> on August 26, 2002.
- Miller, C. S., & Remington, R. W. (2002). Effects of structure and label ambiguity on information navigation. *Conference Extended Abstracts on Human Factors in Computer Systems* (pp. 630 – 631). New York: ACM Press.
- Miller, C. S., & Remington, R. W. (2004). Modeling information navigation: Implications for information architecture. *Human-Computer Interaction*, *19*(3), 225 – 271.
- Miller, D. P. (1981). The depth/breadth tradeoff in hierarchical computer menus. *Proceedings of the Human Factors Society*, Vol. 25 (pp. 296–300).
- Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, *38*(11), 39–41.
- Nielsen, J., & Mack, R. L. (Eds.). (1994). *Usability inspection methods*. Indianapolis: Wiley.
- Pedersen, T., Patwardhan, S., & Michelizzi, J. (2004). Wordnet::similarity - measuring the relatedness of concepts. *Proceedings of Fifth Annual Meeting of the North American Chapter of the Association for Computational Linguistics*.
- Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*, 643 – 675.
- Pirolli, P., & Card, S. (2000). A web site user model should at least predict something about users. *internetworking*, *3*.
- Pirolli, P., & Fu, W.-T. (2003). Snif-act: A model of information foraging on the world wide web. *Proceedings of the Ninth International Conference on User Modeling*.
- Simon, H. A. (1981). *The sciences of the artificial, second edition*. Cambridge, MA: The MIT Press.
- Snowberry, K., Parkinson, S. R., & Sisson, N. (1999). Computer display menus. *Ergonomics*, *26*, 699–712.
- Triesman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology*, *114*(3), 285 – 310.
- Young, R. M. (1998). Rational analysis of exploratory choice. In M. Oaksford, & N. Chater (Eds.), *Rational models of cognition* (pp. 469–500). Oxford University Press.