Information Scent and Web Navigation:
Theory, Models, and Automated Usability Evaluation

Peter Pirolli

PARC
3333 Coyote Hill Rd, Palo Alto, CA 94304, USA
pirolli@parc.com

Wai-tat Fu

Carnegie Mellon University
Department of Psychology
5000 Forbes Ave, Pittsburgh, PA 15213, USA
wfu@cmu.edu

Ed Chi

PARC
3333 Coyote Hill Rd, Palo Alto, CA 94304, USA
echi@parc.com

Ayman Farahat

PARC
3333 Coyote Hill Rd, Palo Alto, CA 94304, USA
farahat@parc.com

Abstract

Within a more Information Foraging Theory, we have developed a rational analysis of Web use, which has shaped a cognitive model of Web navigation called SNIF-ACT. An automated and practical method for initializing the model with requisite knowledge of information scent was developed based on Pointwise Mutual Information (PMI) computations from a local document corpus with a Web back-off. An automated Web usability tool called Bloodhound was developed that implements an algorithm that approximates the operation of the cognitive model. We report on succesful empirical tests of the SNIF-ACT cognitive mode, the PMI method, and Bloodhound.

1 Introduction

Information Foraging Theory (Pirolli & Card, 1999) has been used to develop cognitive models of Web navigation that form the basis for a system that predicts Web usability. Information Foraging Theory assumes that the information-seeking behaviour of users is adaptive within the structure and constraints of the human-information interaction environment in which they work. Users prefer strategies and technologies that maximize the amount valuable information they gain as a function of the interaction cost that they invest. To develop a specific model within Information Foraging Theory involves (a) a *rational analysis* (J. R. Anderson, 1990; Oaksford & Chater, 1998) of the human-information interaction environment, which in turn shapes the specification of (b) *computational cognitive models* implemented as production systems (J. R. Anderson *et al.*, 2004). This paper summarizes a rational analysis of Web navigation that leads to a computational cognitive model called SNIF-ACT (Pirolli & Fu, 2003), which in turn forms the basis of an automated Web usability system called Bloodhound (Chi *et al.*, 2003). A key component of this work is a theory of *information scent* (Pirolli, 2003), which is a psychological theory of how people use perceptual cues, such as World Wide Web (WWW) links in order to make information-seeking decisions and to gain an overall sense of the contents of information collections.

2 Rational Analysis of Web Navigation

The rational analysis approach (J. R. Anderson, 1990; Oaksford & Chater, 1998) involves a kind of reverse engineering in which the theorist asks (a) *what* environmental problem is solved, (b) *why* is a given behavioral strategy a good solution to the problem, and (c) *how* is that solution realized by cognitive mechanism. The products of this approach include (a) characterizations of the relevant goals and environment, (b) mathematical rational choice models (e.g., optimization models) of idealized behavioral strategies for achieving those goals in that environment, and (c) computational cognitive models. This methodology is founded on the heuristic assumption that evolving, behaving systems are well-designed (rational) for fulfilling certain functions in certain environments.

Pirolli’s (in press) rational analyses of information foraging on the Web focused on some of the problems posed by the general task environment faced by Web users, and the structure and constraints of the information environment on the Web. Among these problems were (a) the choice of the most cost-effective and useful browsing actions to take based on the relation of a user’s information need to the perceived proximal cues (information scent) associated with Web links and (b) the decision of whether to continue at a Web site or leave based on ongoing assessments of the site’s potential usefulness and costs. Rational choice models, and specifically approaches borrowed and modified from optimal foraging theory (Stephens and Krebs, 1986) and microeconomics (McFadden, 1974), were used to predict rational behavioral solutions to these problems.

2.1 Link Choice: A Spreading Activation Model of Information Scent

Information foraging behavior will often depend on assessments of the utility and costs of pursuing information items. In browsing for information on the Web, people must base navigation decisions on assessments of *information scent* cues associated with links from one Web page to another. These information scent cues are the small snippets of text and graphics that are associated Web links. Those cues are intended to represent tersely the content that will be encountered by choosing a particular link on one page and navigating to the linked page. When browsing the Web by following links, users must use these cues presented proximally on the Web pages they are currently viewing in order to make navigation decisions. The perceived relevance of the proximal link cues and the distal information they lead to is measured by information scent. If a link cue is perceived to have high information scent, the user will assess that the link is likely to lead to the information goal of the user. The measure of information scent therefore provides a means to predict how users will evaluate different links on a Web page, and as a consequence, the likelihood that a particular link will be followed.

The rational analysis of the use of information scent assumes that the goal of the information forager is to use proximal external information scent cues (e.g., a Web link) to predict the utility of distal sources of content (i.e., the Web page associated with a Web link), and to choose to navigate the links having the maximum expected utility. Pirolli (in press) decomposed this problem into three parts: (1) a Bayesian analysis of the expected relevance of a distal source of content conditional on the available information scent cues, (2) a mapping of this Bayesian model of information scent onto a mathematical formulation of spreading activation, and (3) a model of rational choice that uses spreading activation to evaluate the utility of alternative choices of Web links. This rational analysis yielded a *spreading activation* theory of utility and choice.

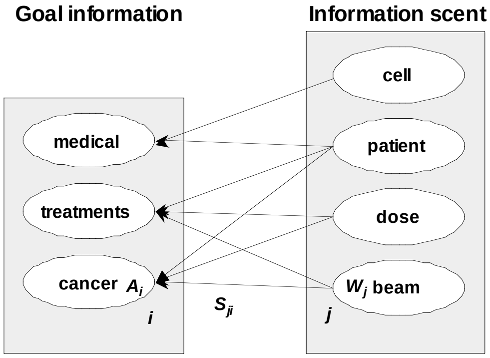

Figure 1. A cognitive structure in which cognitive chunks representing an information goal are associated with chunks representing information scent cues from a Web link.

The spreading activation theory of information scent assumes that the user’s cognitive system represents information scent cues and information goals in cognitive structures called *chunks* Figure 1 presents a schematic example of the information scent assessment subtask facing a Web user. Figure 1 assumes that a user has the goal of finding information about “medical treatments for cancer,” and encounters a Web link labelled with the text that includes “cell”, “patient”, “dose”, and “beam”. The user’s cognitive task is to predict the likelihood that a distal source of content contains desired information based on the proximal information scent cues available in the Web link labels. Each node in Figure 1 represents a cognitive chunk. Chunks representing information scent cues are presented on the right side of Figure 1, chunks representing the user’s information need are presented on the left side. Also represented by lines in Figure 1 are *associations* among the chunks. The associations among chunks come from past experience. The strength of associations reflects the degree to which proximal information scent cues predict the occurrence of unobserved features. The strength of association between a chunk *i* and chunk *j* is computed as,

$$S_{ji} = \log\left(\frac{\Pr(i \mid j)}{\Pr(i)}\right), \tag{1}$$

Where Pr(*i*|*j*) is the probability (based on past experience) that chunk *i* has occurred when chunk *j* has occurred in the environment, and Pr(*i*) is the base rate probability of chunk *i* occurring in the environment. Equation 1 is also known as *Pointwise Mutual Information* (Manning & Schuetze, 1999) or PMI, which is discussed below.

We assume that when a user focuses attention on a Web link their attention to information scent cues activates corresponding cognitive chunks. Activation spreads from those attended chunks along associations to related chunks. For instance, activation would flow from the chunks on the right of Figure 1 through associations to chunks on the left of Figure 1. The amount of activation accumulating on the representation of a user's information goal provides an indicator of the likelihood that a distal source of information has desirable features based on the information scent cues immediately available to the user. For each chunk *i* involved in the user's goal, the accumulated activation received from all associated information scents chunks *j* is,

$$A_i = \sum_j W_j S_{ji}\ , \tag{2}$$

where W_j represents the amount of attention devoted to chunk *j*. The total amount of activation received by all goal chunks *i* is just,

$$V = \sum_i A_i\ . \tag{3}$$

We assume that the utility of choosing a particular link is just the sum of activation it receives (Equation 3) plus some random noise. From this assumption (see Pirolli, in press) we can derive that the probability that a user will choose link *L*, having a summed activation V_L, from a set of links *C* on a Web page, given an information goal, *G*, to be

$$\Pr(L \mid G, C) = \frac{e^{\mu V_L}}{\sum_{k \in C} e^{\mu V_k}}\ . \tag{4}$$

2.2 Patch-leaving Policy

Another decision facing a Web forager is whether to continue navigating a particular Web site or leave. Pirolli's (in press) rational analysis of this problem employs a modified optimal foraging model developed by McNamara (1982). It is assumed that the user employs learning mechanisms to develop an assessment of the potential yield of a Web site, based on the user's current experiential state *x*. This *potential function h*(x) is

$$h(x) = U(x) - C(t) \tag{5}$$

where *U*(x) is the utility of continued foraging in the current Web site (or *information patch*) *x*, and *C*(*t*) is the *opportunity cost* of foraging for the *t* amount of time that is expected to be spent in the information patch. So long as the potential of the Web site is positive (the utility of continuing is greater than the opportunity cost) then the user will continue foraging. The SNIF-ACT model described below incorporates cognitive mechanisms for implementing the judgment to stay or leave based on a version of the potential function in Equation 5.

3 SNIF-ACT

A model called SNIF-ACT (Pirolli & Fu, 2003) was developed based on the measure of information scent and the random utility theory as described above. In this article we present old (SNIF-ACT 1.0) and new data and the newest version of the model (SNIF-ACT 2.0). The basic structure of the SNIF-ACT 1.0 model is shown in Figure 2. Similar to ACT-R models (J. R. Anderson et al., 2004), SNIF-ACT has two memory components – the declarative memory component and the procedural memory component. Elements in the declarative memory component can be contemplated or reflected upon, whereas elements in the procedural memory component are tacit and directly embodied in physical or cognitive activity.

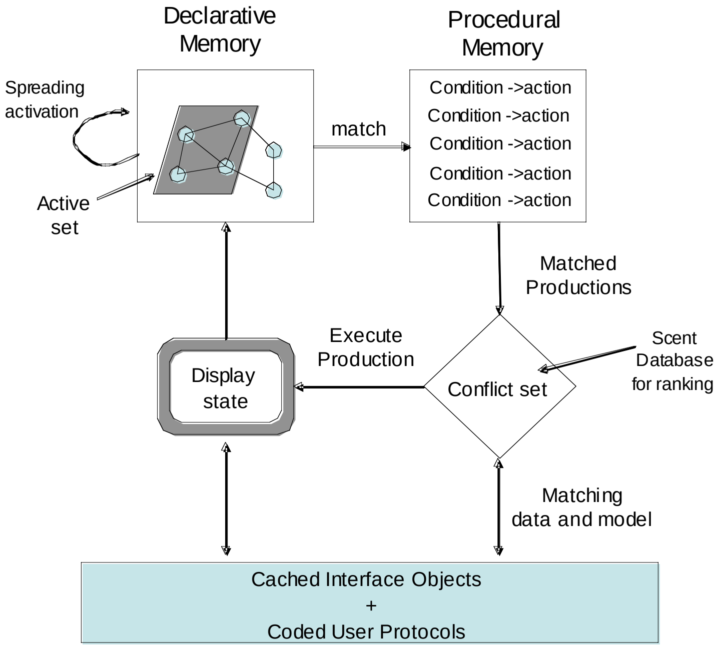

Figure 2. The SNIF-ACT 1.0 architecture.

3.1 Declarative Knowledge

Declarative knowledge in represented as chunks, as depicted in Figure 1, and corresponds to things that we are aware we know and that can be easily described to others, such as the content of Web links, or the functionality of browser buttons, and the current goal of the users (e.g., evaluating a link, choosing a link, etc.). Since our goal is to not to model how users learn to use the browser, we assume that the model has all the knowledge necessary to use the browser, such as clicking on a link, or clicking on the "back" button to go back to the previous Web page. We also assume that users have perfect knowledge of the addresses of most popular Web search engines. Declarative knowledge is simply provided to the model in all the simulations.

3.2 Procedural Knowledge

Procedural knowledge is represented as production rules. For instance, the following is an English gloss of a production rule, Click-link, that is used to simulate the choice of a Web link,

Click-link:
IF the goal is to process a link
& there is a task description
& there is a browser
& there is a link that has been read
& the link has a link description
THEN
Click on the link

If selected, the rule will execute the action of clicking on the link. A production rule has a condition side and an action side. When the all the conditions on the condition side are matched, the production may be fired and when it does, the actions on the action side of the production will be executed. At any point in time, only a single production can fire. When there is more than one match, the matching productions form a "conflict set". One production is then selected from the conflict set based on it's utility based on information scent.

3.3 Selection of Productions

In a SNIF-ACT simulation, information scent cues on a computer display activate chunks and activation spreads through the declarative network of chunks. The amount of activation accumulating on the chunks matched by a production is used to evaluate and select productions. The activation of chunks matched by production rules is used to determine the utility of selecting those production rules. For instance, the utility of the Click-link production described is based on the activation that spreads from the link that it matches against.

3.4 Simulation Results

Two versions of SNIF-ACT have been tested against user data. SNIF-ACT 1.0 was used to predict the behaviour of *N* = 4 Web users studied in detail in Card et al. (2001). SNIF-ACT 2.0 was used to predict the behaviour of N = 244 Web users studied by Chi et al. (2003).

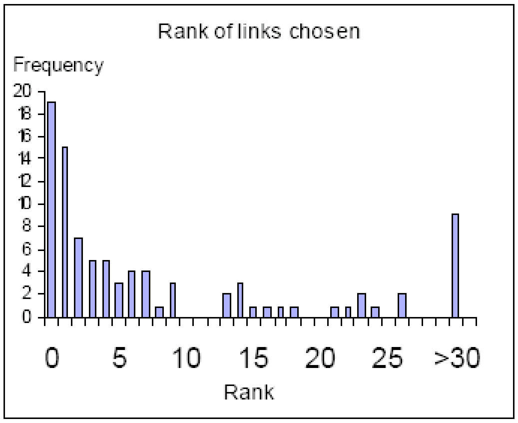

Figure 3. The ranks of the links chosen by participants, as evaluated by SNIF-ACT 1.0. The lower the rank, the more likely that the model will choose the links

3.4.1 SNIF-ACT 1.0

SNIF-ACT 1.0 was used to simulate *N* = 4 users working on two tasks each. Users were free to navigate anywhere on the Web to accomplish these tasks (for details, see Card et al., 2001). Figure 3 comes from the evaluation of the SNIF-ACT 1.0 model by Pirolli and Fu (2003). Figure 3 plots data extracted from all the places where the SNIF-ACT 1.0 simulation was compared against user data at the point when the user was just about to make a selection of a link on a Web page (there are a total of 91 link selections in the data set, which comprise 48% of all the actions performed by four subjects on two tasks). Actions associated with following the links on a page were ranked in SNIF-ACT 1.0 by their information scent utilities, as computed by spreading activation. The x-axis in Figure 3 plots SNIF-ACT 1.0's scent-based ranking of all the possible link-following actions available to the users at each decision point. The y-axis in Figure 3 plots the observed frequency with which the potential SNIF-ACT 1.0 action matched a real user. Figure 3 shows that link choice is strongly related to the information scent values computed in SNIF-ACT 1.0.

SNIF-ACT 1.0 was also used to predict the point at which users leave a Web site. These data are presented in Figure 4. Each data point is the average of *N* = 12 site-leaving actions observed in the data set. The x-axis indexes the four steps made prior to leaving a site (Last-3, Last-2, Last-1, Leave-Site). The y-axis in Figure 4 corresponds to the average information scent value computed by the SNIF-ACT spreading activation mechanisms. The horizontal dotted line indicates the theoretically predicted threshold for leaving a Web site (which in this case is the average information scent value of the page visited by users after they left a Web site). Figure 4 suggests that users essentially assess the expected utility of continuing on at an information patch (i.e., a Web site) against the expected utility of switching their foraging to a new information patch as predicted in Equation 5.

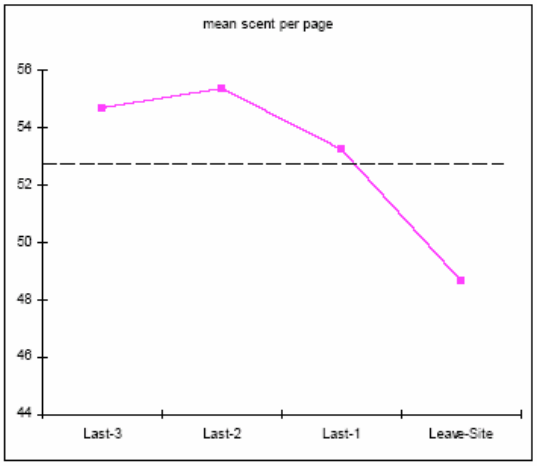

Figure 4. The mean scent scores before participants left a Web site. The dashed line represents the theoretical threshold for leaving a Web site.

3.4.2 SNIF-ACT 2.0

SNIF-ACT 2.0 was matched to data from *N* = 244 users. Users could work on tasks on two Web sites (Yahoo; Parcweb). There were eight tasks for each site. On average, 40 users worked on each of the 8 X 2 = 16 tasks. Users were constrained to never leave the given Web site when performing their given task. The structure of SNIF-ACT 2.0 is similar to that shown in Figure 2, except that the selection of productions is automated by utility calculations in the conflict resolution process (quote). This allows Monte Carlo simulations of the model to generate data for the 16 tasks and matching the frequency that it selected the same links as users. Figure 5 shows that SNIF-ACT 2.0 provides good match to the data.

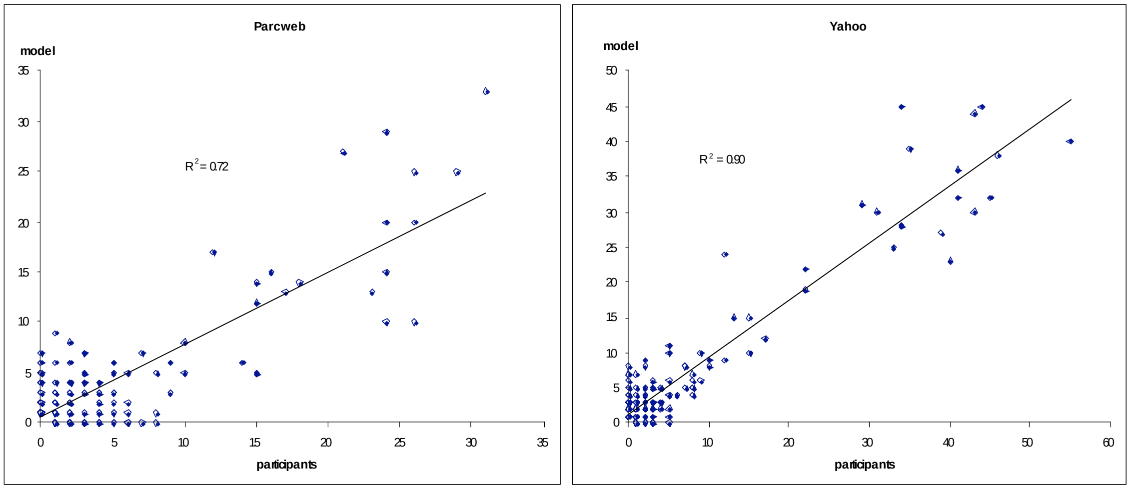

Figure 5. The scatter plots for the number of times links were selected by the model and by user at the Parcweb and Yahoo sites (8 tasks per site).

3.5 Summary

SNIF-ACT is a computational model derived from the rational analyses of Web navigation. The major assumption of the model is that Web navigations can be characterized by mechanisms that maximizes expected information gain. Expected information gain is estimated by a spreading activation mechanism that calculates the relatedness of information goal and link text. Link text that is highly related to the information goal is said to have a high information scent. We showed that the model provides good fits to human link choices in a variety of tasks. The good fits to human data provide strong support for the use of information scent to characterize information-seeking decisions on the Web. We will show later how we capitalize on this finding to build a system that performs automatic usability analysis of Web sites.

4 Computing Strength of Associations Using Pointwise Mututal Information

4.1 Similarity Measures

Conceptually, a measure of information scent captures the word association (or inter-word similarity) of cues on a Web page to a user's information need. Three of the most effective techniques for measuring word association are based on statistical techniques that use corpus co-occurrence counts, latent semantic analysis techniques, or a lexical ontology. However, the throughput and performance of these approaches is limited by search engine limitations or by out of corpus words. We use a novel approach for computing world pair associations. This approach uses Point Wise Mutual Information (PMI) to compute the strength of word associations in Equation 1. PMI is computed from co-occurence counts in a local corpus in case of common terms or from the Web in the case of infrequent terms.

4.1.1 PointWise Mutual Information

Pointwise Mutual Information indicates the amount of information (reduction in uncertainty) about one event that is provided by the observation of another event. In the case of word associations, it indicates the reduction in uncertainty of word "w_2" occurring given that one has observed word "w_1"

To compute PMI, we map each of the two words w_1 , w_2 to the binary variables X, Y respectively. The random variable X takes the value 1 if the word w_1 appears in a particular document and ``0" otherwise. The random variable Y is similarly defined. The PMI is given by

$$SA(w_1\ w_2) \quad = \quad \log \frac{p(X=1,Y=1)}{p(X=1)P(Y=1)} \tag{5}$$

4.1.2 Spreading Activation

The spreading activation framework Pirolli[84] uses a semantic network to model human memory. Anderson (J. R. a. P. L. P. Anderson, 1984) developed a Bayesian analysis of the retrieval problems faced by human memory and argued that one aspect of the optimal memory design would order the retrieval of chunks of memory by an association strength to retrieval cues (i.e., the cues triggering a retrieval from memory). Anderson's Bayesian analysis lead to a specification of association strength that is equivalent to PMI. In spreading activation network, nodes of the network represent memory chunks (roughly, concepts) and the edges represents connection between the chunks. Edges are labeled with association strengths. Anderson's (J. R. a. P. L. P. Anderson, 1984) rational analysis of memory led to a strength of association measure that reflected the log likelihood odds of one event occurring in the context of another:

$$SA(w_1,w_2) = \log \frac{p(X=1,Y=1)}{P(X=1,Y=0)} \tag{6}$$

This strength of association measure is virtually equivalent to PMI for any reasonably sized sample of language. To see this, we rewrite Equation 6 using Bayes law

$$\begin{aligned} p(X=1) &= p(x=1 \mid Y=1) + p(X=1 \mid Y=0) \\ p(X=1) &\approx p(X=1 \mid Y=0) \\ SA(w_1,w_2) &= \log \frac{P(X=1,Y=1)}{P(Y=1)} \end{aligned} \tag{7}$$

where Equation 7 follows from the fact that the probability of observing a word p(w)is extremely small to the probability of not observing the word. Equation 7 is the basis for the strength of association measure in Equation 1.

4.2 System Description and Evaluation

We developed a hybrid a system for computing co-occurrence counts from a sample corpus. Our system is based on the Lucene search engine. We used Lucene to index the first 10 million pages of the Stanford Webbase project. The hybrid system used the local crawl to compute the co-occurrence counts and backed of to the Web in case of low counts. We used the search engines Google and AltaVista to compute the Web co-occurrence counts.

We evaluated our approach on two datasets. The first set was the synonym test from Test of English as a Foreign Language TOEFL data (Landauer & Dumais, 1997). This set consisted of 80 problem words. For each problem word, we are given four alternative words, which of the alternatives is most similar in meaning to the problem word? The second data set was the synonym test from the Readers Digest Data. This set consisted of 100 words problem and four alternatives for each problem.

The Hybrid approach as shown in Table 1 outperforms the Web based approach. In addition, by limiting the search engine queries, we can achieve high throughputs. We can currently compute the pair wise similarities between 40,000 terms in less than 16 hours.

Table 1. Evaluations of variations on the PMI method.

Measure	Hybrid	Local	Web	Lift	Miss	Savings
TOEFL	53	51	46	15%	18	94%
Reader	60	52	56	7%	180	55%

5 Bloodhound

5.1 The Service

The Bloodhound service (Chi et al., 2003) employs a Web user flow model to predict Web site usage patterns and identify Web site navigation problems. The service employs a variation on the WUFIS (Web User Flow by Information Scent) algorithm which abstracts away from the details of the SNIF-ACT model. This assumes that users come to a Web site with some information goal and forage for information by choosing links based on proximal information scent cues.

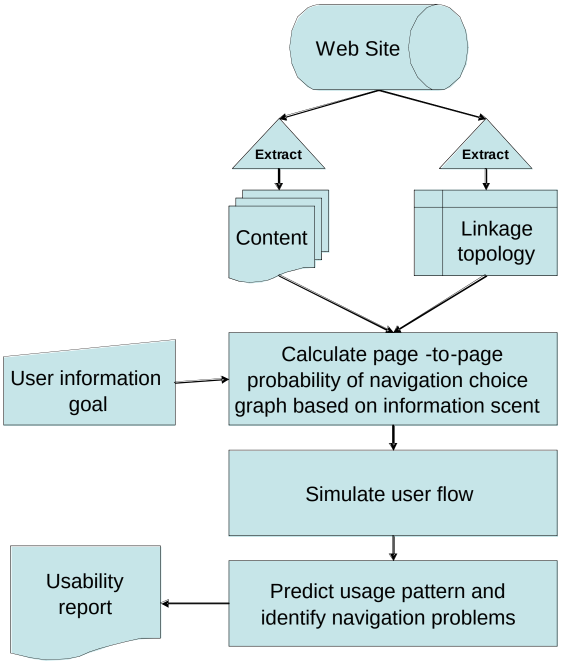

Figure 6. The conceptual flow chart for the processing done by the Bloodhound Web usability service.

Figure 6 presents an overview of the process involved used by the Bloodhound service. A person (the *Web site analyst*) interested in performing a usability analysis of a Web site must indicate the Web site to be analyzed, and provide a candidate user information goal representing a task that users are expected to be performing at the site. Bloodhound then must crawl the Web site to develop a representation of the linkage topology (the page-to-page links) and download the Web pages (content). From these data, Bloodhound analyzes the Web pages to determine the proximal information scent cues associated with every link on every page. At this point Bloodhound essentially has a representation of every page-to-page link, and the proximal cues associated with that link. From this, Bloodhound develops a graph representation in which the nodes are the Web site pages, the vertices are the page-to-page links at the site, and weights on the vertices represent the probability of a user choosing a particular vertex given the user's information goal and the proximal information scent cues associated with the link (e.g., Equation 4). This graph is represented as a page-by-page matrix in which the rows represent individual unique pages at the site, the columns also represent Web site pages, and the matrix cells contain the navigation choice probabilities that predict the probability that a user with the given information goal, at a given page, will choose to go to a linked page. Using matrix computations, this matrix is used to simulate user flow at the Web site by assuming that the user starts at some given Web page and iteratively chooses to go to new pages based on the predicted navigation choice probabilities. The user flow simulation yields predictions concerning the pattern of visits to Web pages, and the proportion of users that will arrive at target Web pages contain the information relevant to their tasks.

The Bloodhound service is provided over the Web. An input screen is provided to Web site analysts that allows them to enter specifications of user tasks, the Web site URL, and the target pages that contain the information relevant to those tasks. An analysis is then performed by Bloodhound and report is then automatically generated that indicates such things as the predicted number of users who will be able to find target information relevant to the specified task, and intermediate navigation pages that are predicted to be highly visited that may be a cause of bottlenecks.

5.2 Evaluation

Chi et al. (2003) performed an evaluation of the capability of Bloodhound to predict actual user navigation patterns. Users were solicited to perform Web tasks at home, office, or place of their choosing and their performance was logged using a remote usability testing system. A total of $N = 244$ users participated in the study. Four different types of Web sites were studied with eight tasks of varying difficulty for each site. The comparison of interest was the match between observed and predicted usage patterns for each task and Web site. For each task + Web site, the observed data were the distribution of the frequency of page visits over every Web page. For instance, for a particular task + Web site, the home page might be visited 75 times, another page 25 times, and so on. The comparison was the distribution of page visits for that task and Web site as predicted by Bloodhoound. Of the 4 X 8 = 32 combinations of Web sites and tasks, there were strong correlations (Pearson $r > 0.8$) of observed and predicted visitation frequencies for twelve cases, moderate correlations ($0.5 \leq r \leq 0.8$) for seventeen cases, and weak correlations ($r < 0.5$) for three cases. Given that this was the first evaluation of Bloodhound the results seemed like a validation of the promise of the approach.

6 General Discussion

Within a more Information Foraging Theory, we have developed a rational analysis of Web use, which has shaped a cognitive model of Web navigation called SNIF-ACT. An automated and practical method for initializing the model with requisite knowledge of information scent was developed based on PMI computations from a local document corpus with a Web back-off. An automated Web usability tool called Bloodhound was developed that implements an algorithm that approximates the operation of the cognitive model.

In general, this work has many commonalities with CWW (Blackmon *et al.,* 2002) and MESA (Miller & Remington, 2004). In all of these systems, the user is modeled as an agent who searches through a space of decision states, corresponding to Web pages, at which the user is faced with a set of alternative actions to choose, and the alternatives are evaluated by some version of information scent. In essence the user is modeled as performing a kind of heuristic hill-climbing search, where information scent provides the heuristic. These models have been tested against data from user performing tasks that are novel (unfamiliar) where such heuristic search would be expected. One question, for all these models, is how well they can be extended to modeling tasks in which the users have considerable background knowledge or expertise.

Acknowledgements

Portions of this research have been supported by an Office of Naval Research Contract No. N00014-96-C-0097 to P. Pirolli and S.K. Card and Advanced Research and Development Activity, Novel Intelligence from Massive Data Program Contract No. MDA904-03-C-0404 to S.K. Card and Peter Pirolli.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review, 11*(4), 1036-1060.

Anderson, J. R. a. P. L. P. (1984). Spread of activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 791-798.

Blackmon, M. H., Polson, P. G., Kitajima, M., & Lewis, C. (2002). Cognitive walkthrough for the web. *CHI 2002, ACM Conference on Human Factors in Computing Systems, CHI Letters, 4*(1).

Card, S., Pirolli, P., Van Der Wege, M., Morrison, J., Reeder, R., Schraedley, P., et al. (2001). Information scent as a driver of web behavior graphs: Results of a protocol analysis method for web usability. *CHI 2001, ACM Conference on Human Factors in Computing Systems, CHI Letters, 3*(1), 498-505.

Chi, E. H., Rosien, A., Suppattanasiri, G., Williams, A., Royer, C., Chow, C., et al. (2003). The bloodhound project: Automating discovery of web usability issues using the infoscent simulator. *CHI 2003, ACM Conference on Human Factors in Computing Systems, CHI Letters, 5*(1), 505-512.

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.

Manning, C. D., & Schuetze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

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

Oaksford, M., & Chater, N. (Eds.). (1998). *Rational models of cognition*. Oxford: Oxford University Press.

Pirolli, P. (2003). A theory of information scent. In J. Jacko & C. Stephanidis (Eds.), *Human-computer interaction* (Vol. 1, pp. 213-217). Mahwah, NJ: Lawrence Erlbaum.

Pirolli, P., & Card, S. K. (1999). Information foraging. *Psychological Review, 106*, 643-675.

Pirolli, P., & Fu, W. (2003). Snif-act: A model of information foraging on the world wide web. In P. Brusilovsky, A. Corbett & F. de Rosis (Eds.), *User modeling 2003, 9th international conference, um 2003* (Vol. 2702, pp. 45-54). Johnstown, PA: Springer-Verlag.