SNIF-ACT: A Model of Information Foraging on the World Wide Web

Peter Pirolli and Wai-Tat Fu[1]

PARC, 3333 Coyote Hill Road, Palo Alto, California 94304
pirolli@parc.com, wfu@gmu.edu

Abstract. SNIF-ACT (Scent-based Navigation and Information Foraging in the ACT architecture) has been developed to simulate users as they perform unfamiliar information-seeking tasks on the World Wide Web (WWW). SNIF-ACT selects actions based on the measure of *information scent*, which is calculated by a spreading activation mechanism that captures the mutual relevance of the contents of a WWW page to the goal of the user. There are two main predictions of SNIF-ACT: (1) users working on unfamiliar tasks are expected to choose links that have high information scent, (2) users will leave a site when the information scent of the site diminishes below a certain threshold. SNIF-ACT produced good fits to data collected from four users working on two tasks each. The results suggest that the current content-based spreading activation SNIF-ACT model is able to generate useful predictions about complex user-WWW interactions.

1 Introduction

Over the course of the past decade and a half, vast amounts of content in the form of hypermedia have become available to the average computer user, primarily through the World Wide Web (WWW). Over this time, there has been limited progress towards a deep scientific understanding of the psychology of human interaction with the WWW. Detailed cognitive models are difficult to create, and unlike most models in human-computer interaction, the analysis of human-WWW interaction requires modeling user interaction with the semantics of Web content. Although difficult, there are many potential payoffs to developing a scientific foundation in this area. Such cognitive models could provide insights and engineering principles for improving usability. More directly they could provide automated cognitive engineering tools. They could serve as the basis for user models embedded in systems and devices to improve interaction, and they could serve as the basis for helping people to learn how to find, make sense of, and use information to improve solutions to significant everyday problems involving health, finance, career, and so on.

The purpose of this paper is to present a computational cognitive model that addresses data collected from WWW users studied in a laboratory setting using realistic tasks. The model is based on a theoretical integration of *Information Foraging Theory* [1] and ACT-R [2]. Particularly important is the concept of *information scent*, which characterizes how users evaluate the utility of hypermedia actions. SNIF-ACT has been developed with a *user-trace methodology* [3] for studying and analyzing the psychology of users performing ecologically valid WWW tasks. A *user trace* is a record of all significant states and events in the user-WWW interaction based on eye tracking data, application-level logs, and think-aloud protocols. A *user-tracing architecture* has been implemented for developing simulation models of user-WWW interaction and for comparing SNIF-ACT simulations against user-trace data. The user tracing architecture compares each action of the SNIF-ACT simulation directly against observed user actions.

Previous cognitive models of HCI have mainly been developed to deal with the analysis of expert performance on well-defined tasks involving application programs [4]. These have had limited applicability to understanding foraging through content-rich hypermedia, and consequently new theories are needed. An attempt [5] at developing a GOMS model [6] of WWW users failed to have a significant correlation with user behavior [7]. That model's behavior was based purely on the structure of pages and links—none of the behavior was determined by the semantics of

---

[1] Present address: George Mason University, MS 3e5 4400 University Drive, Fairfax, VA 22030.

page content. A day-in-the-life protocol analysis of real users engaged in their own tasks with the WWW [8] showed that the majority of user time was devoted to processing WWW content. Recent work [9] involving comprehension-based models of exploration (that use Latent Semantic Analysis) has achieved some promising success in modeling limited WWW interaction. These results lead us to conclude that a realistic user model of WWW interaction will have to deal with how user behavior depends on content.

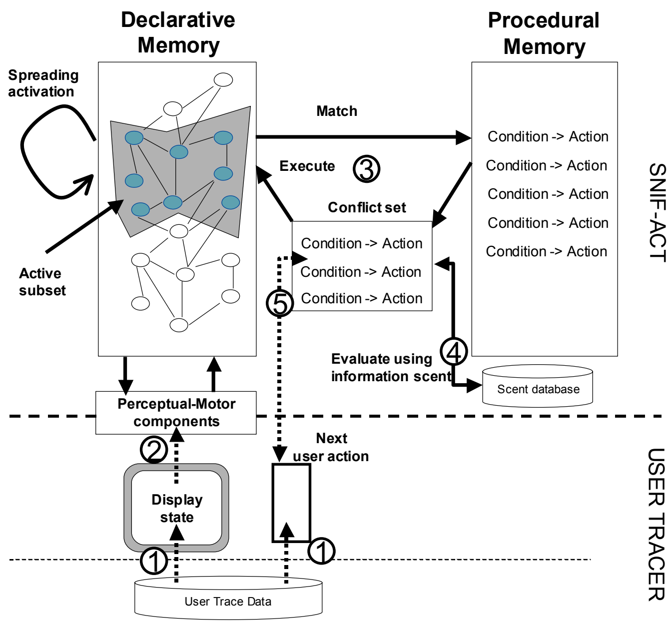

Fig. 1. The structure of the SNIF-ACT model and the User-Tracing Architecture. The numbers indicate the order of the steps in each cycle of the SNIF-ACT simulation.

2 SNIF-ACT

The goal of our modeling effort is to develop a computer program than simulates the user in enough detail to reproduce the user data. SNIF-ACT (Figure 1) is the model that we are currently developing to simulate WWW users. SNIF-ACT is an extension of the ACT-R theory and simulation environment [2], a general *production system* architecture designed to model human psychology. By using this system to model WWW behavior, we link our analysis to the same principles used to model cognitive behavior in general. ACT-R contains principles concerning: (1) knowledge representation, (2) knowledge deployment (performance), and (3) knowledge acquisition (learning). There are two major components in the ACT-R architecture: a *declarative knowledge* component and a *procedural knowledge* component. ACT-R has two kinds of memory for these two different kinds of knowledge.

2.1 Declarative Knowledge

Declarative knowledge corresponds to things that we are aware we know and that can be easily described to others, such as the content of WWW links, or the functionality of browser buttons. Declarative knowledge is represented formally as *chunks* in ACT-R. Declarative chunks in ACT-R have sub-symbolic *activation*. Activation represents the log-odds of how likely a piece of knowledge is needed at a particular time, and may be interpreted metaphorically as a kind of mental energy that drives cognitive processing. Activation spreads from the current focus of attention, including goals, through *associations* among chunks in declarative memory. These associations are built up from experience, and they reflect how ideas co-occur in cognitive processing. Generally, activation-based theories of memory predict that more activated knowledge structures will receive more favorable processing. Chunks with higher activation values take less time to use and have a greater chance to have an impact on behavior. Activation is a way of quantifying the degree of relevance of declarative information to the current focus of attention (mathematically, it represents the posterior probability of how likely each piece of declarative information is needed given the current focus of attention). At any point in time, there is a stack of goals encoding the user's intentions. Goals are also represented as chunks. ACT-R is always trying to achieve the goal that is on top of that stack, and at any point in time, it is focused on a single goal.

Table 1. Productions in SNIF-ACT. Productions in the same row have the same goal.

Use-Search-Engine: IF the goal is Goal*Start-Next-Patch & there is a task description & there is a browser & there is a search engine name in memory & the browser is not at the search engine THEN Set and push a subgoal Goal*Use-Search-Engine to the goal stack	Type-URL-to-Go-to-Site: IF the goal is Goal*Start-Next-Patch & there is a task description & there is a browser & there is a site (URL) in memory & the browser is not at the site (URL) THEN Set and push a subgoal Goal*Go-To-Site to the goal stack & note that the site is the URL
Go-To-Search-Engine: IF the goal is Goal*Use-Search-Engine & there is a task description & there is a browser & there is a search engine name in memory & the browser is NOT at a search engine THEN Set and push a subgoal Goal*Go-To-Site to the goal stack & note that the site is the search engine	Search-Using-Search-Engine: IF the goal is Goal*Use-Search-Engine & there is a task description & there is a browser & the browser is at a search engine & there are search terms in memory THEN Enter search terms in search engine & pop the current goal from the goal stack
Go-To-Site-By-Bookmark: IF the goal is Goal*Go-To-Site & there is a task description & there is a browser & there is a site in memory & the browser is not at the site THEN Use bookmark to go to site & pop the current goal from the goal stack	Go-To-Site-By-Typing: IF the goal is Goal*Go-To-Site & there is a task description & there is a browser & there is a site in memory & the browser is not at the site THEN Type the URL to go to site & pop the current goal from the goal stack
Start-Process-Page: IF the goal is Goal*Start-Next-Patch & there is a task description & there is a browser & the browser is on an unprocessed page THEN Set & push a subgoal Goal*Process-Page to the goal stack	
Process-Links-on-Page: IF the goal is Goal* Process-Page & there is a task description & there is a browser & there is an unprocessed link THEN Set and push a subgoal Goal*Process-Link to the goal stack	
Attend-to-Link: IF the goal is Goal* Process-Link & there is a task description & there is a browser & there is an unattended link THEN Choose an unattended link and attend to it	Other productions with identical conditions on the left-hand-side (i.e IF side), but different actions (i.e. THEN side). Search-Site-using-Search-Box Backup-a-Page Forward-a-Page Modify-URL Use-Back-History
Read-and-Evaluate-Link: IF the goal is Goal* Process-Link & there is a task description & there is a browser & the current attention is on a link THEN Read and Evaluate the link	
Click-Link: IF the goal is Goal* Process-Link & there is a task description & there is a browser & there is an evaluated link & the link has the highest activation THEN Click on the link	Leave-Site: IF the goal is Goal* Process-Link & there is a task description & there is a browser & there is an evaluated link & the mean activation on page is low THEN Leave the site & pop the goal from the goal stack

2.2 Procedural Knowledge

Procedural knowledge is knowledge (skill) that we display in our behavior without conscious awareness, such as knowledge of how to ride a bike, or how to point a mouse to a menu item. Procedural knowledge specifies how declarative knowledge is transformed into active behavior. Procedural knowledge is represented as condition-action pairs, or *production rules* (Figure 1). For instance, our SNIF-ACT simulation contains the production rule *Use-Search-Engine* (see Table 1). The production applies in situations where the user has a goal to go to a WWW site, has processed a task description, and has a browser in front of them. The production rule specifies that a subgoal will be set to use a search engine. The condition (*IF*) side of the production rule is matched to the current goal and the active chunks in declarative memory, and when a match is found, the action (*THEN*) side of the production rule will be executed. Table 1 also shows most of the productions in the SNIF-ACT simulations, presented in their English-equivalent forms. Productions in the same row have the same goal, and may compete against each other during the simulations.

Roughly, the idea is that each elemental step of cognition corresponds to a production. At any point in time, a single production fires. When there is more than one match, the matching rules form a *conflict set*, and a mechanism called *conflict resolution* is used to decide which production to execute (see Figure 1). The conflict resolution mechanism is based on a utility function. The expected utility of each matching production is calculated based on this utility function, and the one with the highest expected utility will be picked. In modeling WWW users, the utility function is provided by information foraging theory, and specifically the notion of *information scent* [1]. This constitutes a major extension of the ACT-R theory and is described in greater detail below.

2.3 Utility: Information Scent

As users browse the WWW, they make judgments about the utility of different courses of action available to them. Typically, they must use local cues, such as link images and text, to make navigation decisions. Information scent refers to the local cues that users process in making such judgments. The analogy is to organisms that use local smell cues to make judgments about where to go next (for instance in pursuing some prey). In earlier work [1, 10] we extended ACT-R to produce a theory called ACT-IF (where IF stands for "information foraging"). ACT-IF included a formal model of information scent that predicted how users would use text presented in browsers to make navigation decisions. The model of users' judgments of information scent is based on spreading activation. The basic idea is that a user's information goal activates a set of chunks in a user's memory, and text on the display screen activates another set of chunks. Activation spreads from these chunks to related chunks in a *spreading activation network*. Through this spreading activation network, the amount of activation accumulating on the goal chunks and display chunks is an indicator of their mutual relevance. The spreading activation network is therefore content-based, as mutual relevance of user goals and contents are calculated each time the display changes. The amount of activation is used to evaluate and select productions. The activation of content-dependent chunks matched by production rules can be used to determine the utility of selecting those production rules dynamically.

If, for example, the production *Click-Link* is selected, the rule will execute the action of clicking on the link. The chunks associated with the task description and the link description will have a certain amount of activation. That combined activation will be used to evaluate the rule. If there are two Click-link productions matching against chunks for two different links, then the one with more highly activated chunks will be selected. As we describe next, the activation level will tend to reflect the degree of relevance of the link text to the task description.

The spread of activation from one cognitive structure to another is determined by weighting values on the associations among chunks. These weights determine the rate of activation flow among chunks. In the context of WWW browsing, we assume that activation spreads from the user's goal, which is the focus of attention, through memory associations to words and images that the user sees on WWW pages. Associations have *strengths* or *weights* that determine the amount of activation that flows from one chunk to another. If the user reads some link text on a WWW page, and the link text is strongly associated to the user's goal, then we expect the user to judge the link as being highly relevant to the goal.

The association strengths among words in human memory are assumed to be related to the probabilities of word occurrences and of word co-occurrences. Consequently, the spreading activation computation of information scent in SNIF-ACT requires these estimates. In past research [1, 10], we derived these estimates from the Tipster corpus [11]. This database contained statistics relevant to setting the base-level activations of 200 million word tokens and the inter-word association strengths of 55 million word pairs. Unfortunately, the Tipster corpus does not contain many of the novel words that arise in popular media such as the WWW. For instance, the movie title "Antz" does not occur in the Tipster corpus. Consequently, we augment the statistical database derived from Tipster by estimating word frequency and word co-occurrence statistics from the WWW itself using a program that calls on the AltaVista search engine to provide data. As indicated in Figure 1, the spreading activation networks needed to perform the scent computations are stored in a *scent database* that is accessed when production evaluations are computed by SNIF-ACT.

2.4 Predictions

There are two main predictions of the SNIF-ACT model that derive from the utility predictions of the information scent computations, and from the patch model of Information Foraging Theory [1]. The first main prediction concerns link-following behavior. Novice users working on unfamiliar tasks are expected to choose links that have high information scent. Basically, users will be performing a kind of hill-climbing with information scent as the heuristic for choosing the next step to take.

The second prediction concerns the points at which users will give up on WWW sites. A user's information environment has a patchy structure, with information collected together on bookshelves, piles on desks, in folders on personal computers, libraries, etc. Information is encountered at a denser rate when a users is "in" a patch (e.g., viewing a set of WWW search results) as opposed to "between" patches (e.g., formulating a search query and waiting on search results). WWW sites are considered information patches. The conventional model of foraging in information patches predicts that users will leave an information patch when the expected utility of the site diminishes below the utility of moving to another site.

3 User Tracing Method

In previous work [3] we presented a *user-tracing method* aimed at extracting and validating information at an individual user level. Controlled laboratory experiments were conducted [12] using tasks derived from a task database collected by survey from WWW users [13]. The laboratory experiments collected data using an eye tracker, logging software that collects all user interactions with a WWW browser, and video recordings of think-aloud verbal protocols [12]. These data were coded by automatic means and by hand into a comprehensive trace of states and events representing the interaction of user with the WWW. SNIF-ACT models of user cognition and perception were then developed to simulate—as accurately as possible—the observed user-WWW interactions. Data from the users and tasks analyzed by Card et al. [12] were simulated by SNIF-ACT to produce the model fits discussed below.

3.1 Tasks and Users

Tasks for our study were selected from a database collected by a survey of over 2000 WWW users [13]. The two analyzed in details are:

> <u>Antz Task</u>: *After installing a state of the art entertainment center in your den and replacing the furniture and carpeting, your redecorating is almost complete. All that remains to be done is to purchase a set of movie posters to hang on the walls. Find a site where you can purchase the set of four Antz movie posters depicting the princess, the hero, the best friend, and the general.*
>
> <u>City Task:</u> *You are the Chair of Comedic events for Louisiana State University in Baton Rouge, LA. Your computer has just crashed and you have lost several advertisements for upcoming events. You know that The Second City tour is coming to your theater in the spring, but you do not know the precise date. Find the date the comedy troupe is playing on your campus. Also find a photograph of the group to put on the advertisement.*

Four users were solicited from PARC and Stanford. Users were encouraged to perform both tasks as they would typically, but they were also instructed to think out loud [14] as they performed their tasks.

3.2 User Tracing Instrumentation

User trace data consists of several kinds of data recorded and analyzed by our instrumentation package. Performance on the tasks was recorded using an instrumentation package that included: (a) WebLogger [15], which is a program that tracks user keystrokes, mouse-movements, button use, and browser actions, (b) an eye tracker, and (c) video recordings that focused on the screen display. Details of the instrumentation used are given in [12]. WebLogger also saves the actual Web content (i.e. the text, images, scripts, etc.) that a user looked at during a browsing session. It does this by saving a cache of all pages and associated content that was viewed by the user. Eye-movements are handled by our WebEyeMapper system. Videotapes of users thinking aloud provide additional data about users' goals and subgoals, attention, and information representation [14]. The video plus WebLogger and WebEyeMapper data are used to produce a *Web Protocol Transcript*. The Web Protocol Transcript includes interactions recorded by the WebLogger, transcribed audio/video data, and *model coding* of the inferred cognitive action that is associated with the data. The protocol analysis provides data that are not available from WebLogger and WebEyeMapper, especially the users' reading and evaluation of content and links.

3.3 User Trace Comparator

Figure 1 shows how the User Trace Comparator controls the SNIF-ACT simulation model and matches the simulation behavior to the user trace data (each step is indicated by a circle in Figure 1):

1. Parse the User Trace Database to determine the next display state and the next user action that occurs at that display state.
2. If the display state has changed, then indicate this to the SNIF-ACT system. SNIF-ACT contains production rules that actively perceive the display state and update declarative memory to contain chunks that represent the perceived portions of the display.
3. Run SNIF-ACT so that it runs spreading activation to identify the active portion of declarative memory and matches productions against working memory to select a conflict set of production rules.
4. SNIF-ACT evaluates the productions in the conflict set using the information scent computations. At the end of this step, one of the rules in the conflict set will be identified as the production to execute.
5. Compare the production just selected by SNIF-ACT to the next user action and record any statistics (notably whether or not the production and action matched). If there is a match, then execute the production selected by SNIF-ACT. If there is a mismatch, then select and execute the production that matches the user action.
6. Repeat Steps 1 - 5 until there are no more user actions.

4 Results

Figure 2 shows examples of behavior extracted from the two tasks performed by one of the four study participants. The behavior is plotted as a Web Behavior Graph (WBG), which is a version of a problem behavior graph (Newell and Simon, 1972). Each box in the diagram represents a state in a problem space. Each arrow depicts the execution of an operator, moving the state to a new state. Double vertical arrows indicate the return to a previous state, augmented by the experience of having explored the consequences of some possible moves. Thus time in the diagram proceeds left to right and top to bottom. Color surrounding the boxes in Figure 2 represent different WWW sites. An **X** following a node indicates that the user exceeded the time limits for the task and that it was therefore a failure. The WBG in Figure 2, and the WBGs for the remaining study participants and users, is presented in greater detail elsewhere [12]. The WBG is particularly good at showing the structure of the search. One may characterize task difficulty in terms of the branchiness of the WBGs, with more branches indicating that search paths were abandoned and the user returned to a prior state. Another way of characterizing task difficulty is by the number of states visited by users. From Figure 2 it is evident that the ANTZ task is more difficult than the CITY task. This was true over all four users [12]. Since users were all novices in the domain, the difference of the WBGs between the two tasks may indicate how well the hill-climbing heuristics could successfully lead the users to the goals. The goal in this paper was to evaluate how well the SNIF-ACT simulation matched user behavior, such as that depicted in Figure 2, and to assess how much of the variability of the WWW behavior is predictable from information scent.

The predictions made by the SNIF-ACT model were tested against the log files of all data sets. The major controlling variable in the model is the measure of information scent, which predicts two major kinds of actions (Table 2):

(1) which links on a web page people will click on, and (2) when people decide to leave a site. These kinds of actions were therefore extracted from the log files and compared to the predictions made by the model. We call the first kind of actions *link-following* actions, which were logged whenever a participant clicked on a link on a web page. The second kind of actions was called *site-leaving* actions, which were logged whenever a participant left a web site (and went to a different search engine or web site). The two kinds of actions made up 72 % (48% for link-following and 24% for site-leaving actions) of all the 189 actions extracted from the log files.

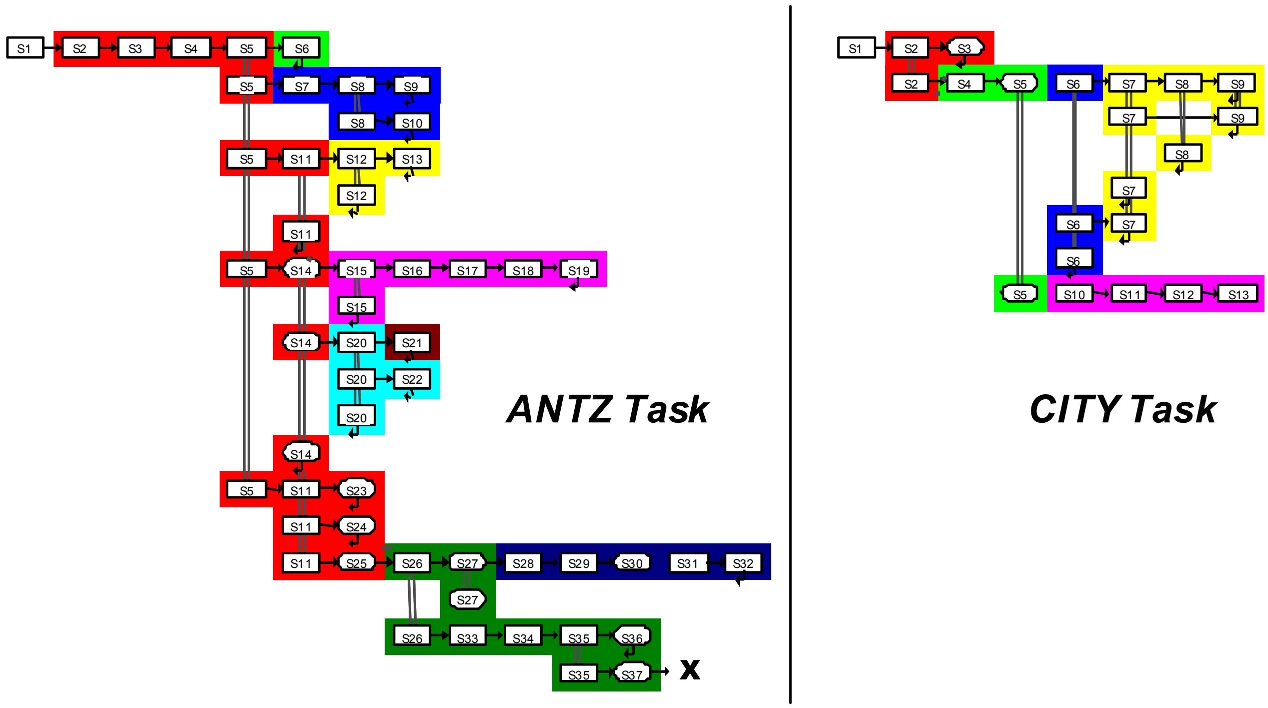

Fig. 2. Web Behavior Graphs for one study participant working on the ANTZ task (left) and CITY task (right).

4.1 Link-following actions

The SNIF-ACT model was matched to the link-following actions extracted from the $N = 8$ (2 tasks x 4 participants) data sets. The user trace comparator was used to compare each action from each participant to the action chosen by the model. Whenever a link-following action was encountered, the SNIF-ACT model ranked all links on the web page according to the information scent of the links. We then compared the links chosen by the participants to the predicted link rankings of the SNIF-ACT model. If there were a purely deterministic relationship between predicted information scent and link choice, then all users would be predicted to choose the highest ranked link. However, we assume that the scent-based utilities are stochastic [16, 17] and subject to some amount of variability due to users and context (which is also consistent with ACT-R [2]). Consequently we expect the probability of link choice to be highest for the links ranked with the greatest amount of scent-based utility, and that link choice probability is expected to decrease for links ranked lower on the basis of their scent-based utility values.

Fig. 3 shows that link choice is strongly related to scent-based utility values. Links ranked higher on scent-based utilities tend to get chosen over links ranked lower. There are a total of 91 link-following actions in Figure 2. The distribution of the predicted link selection was significantly different from random selection $\chi2$ (30) = 18589.45, $p < 0.0001$. This result replicates a similar analysis made by Pirolli and Card [1] concerning the ACT-IF model prediction of cluster selection in the Scatter/Gather browser.

4.2 Site-leaving actions

To test how well information scent is able to predict when people will leave a site, site-leaving actions were extracted from the log files and analyzed. Site-leaving actions are defined as actions that led to a different site (e.g. when the participants used a different search engine, typed in a different url to go to a different web site, etc.) The results were plotted in Fig. 4. Each point in is the mean of the 12 site-leaving actions of the mean information scent of all the links on the page. It shows the four web pages the participants visited before they left the site (i.e. Last-3, Last-2, Last-1, and Leave-Site in Fig. 4). It shows that initially the mean information scent of the web page was high, and right before the participants left the site, the mean information scent dropped.

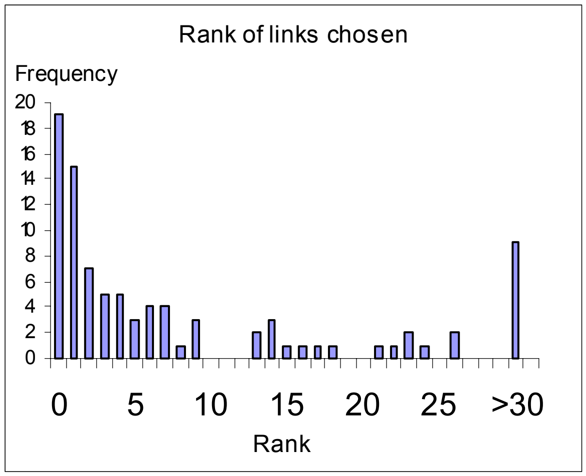

Fig. 3. Frequency that SNIF-ACT productions match link-following actions. The SNIF-ACT production rankings are computed at each simulation cycle over all links on the same web page and all productions that match.

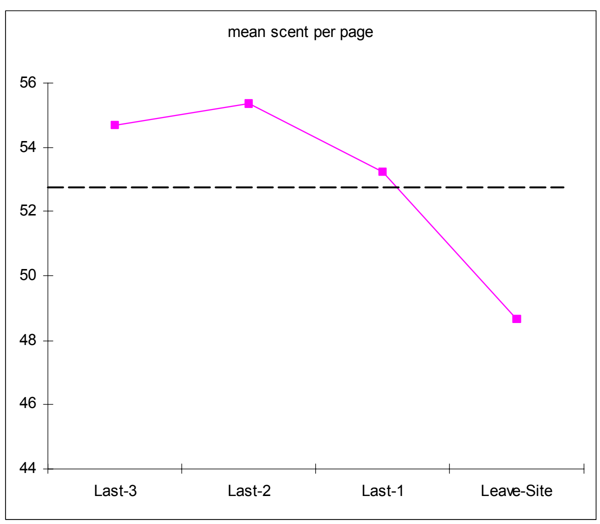

Fig. 4. Mean information scent on the web page as a function of distance from the last web page before the users left the site. The dotted line represent the mean scent of the next site the participants went to after they left the current site.

Fig. 4 also shows the mean information scent of the web pages right after the participants left the site (the dotted line in Fig. 4). It shows that the mean information scent on the page right after they left the site tended to be higher than the mean information scent before they left the site. This is consistent with the information foraging theory which states that people may switch to another "information patch" when the expected gain of searching in the current patch is lower than the expected gain of searching for a new information patch. In fact, from the verbal protocols, we often found utterances like "it seems that I don't have much luck with this site", or "maybe I should try another search engine" right before participants switch to another site. It suggests that the drop in information scent on the web page could be the factor the triggered participants' decision to switch to another site.

4.3 Summary of results

The results show that the measure of information scent is able to generate good predictions to user-WWW interaction. Most of the links chosen by the participants were ranked high by SNIF-ACT, suggesting that SNIF-ACT is able to predict which links people will click on a web page using the measure of information scent. Information scent was also shown to be able to predict when people will leave a site. It was shown that when participants left a site, the average information scent of the site was diminishing. The interesting finding was that the information scent of the web page right before the participants left the site was much lower than that of the site that they switched to. It is possible that, from experience, people have built up an expectation of the information scent value of various web sites. When the information scent value of a particular web site dropped to a value that is below the expected information scent value of other sites, people may decide to abandon the current site and switch to another site.

5 General Discussion

We have described the user modeling and tracing techniques using SNIF-ACT. The system demonstrates how cognitive models can be used to understand real-world user-WWW interactions. We have shown that for general information-seeking tasks in unfamiliar domains, user-WWW interactions depend heavily on the content of the WWW pages. Using the measure of information scent, mutual relevance between the user goals and web contents were captured and stored in scent information databases. The use of information scent, together with the ACT cognitive architecture was shown to predict user actions well. We believe the SNIF-ACT system has provided a good user model of the participants.

Since the participants were seeking information in unfamiliar domains, the current productions in SNIF-ACT represents only weak-method (general) WWW problem solving behavior. Bhanvnani [18] has found that there are expert-novice differences in WWW behavior—for instance, medical experts exhibit strategic knowledge (in addition to weak methods like hill-climbing) in searching for medical information that reduces the branchiness of their WBGs as compared to regular non-expert users. The simple set of productions used in the current SNIF-ACT can be considered the basic knowledge possessed by novice users of the WWW. Specific productions can be added to the system to model the domain-specific knowledge of expert users.

6 Acknowledgements

This work is supported by an Office of Naval Research grant No. N00014-96-C-0097 to P. Pirolli and S. Card. We thank Ed Chi for assistance in generating the Scent Databases for SNIF-ACT.

REFERENCES

1. Pirolli, P. and S.K. Card, *Information foraging.* Psychological Review, 1999. **106**: p. 643-675.
2. Anderson, J.R. and C. Lebiere, *The atomic components of thought*. 2000, Mahwah, NJ: Lawrence Erlbaum Associates.
3. Pirolli, P., et al. *A user-tracing architecture for modeling interaction with the World Wide Web*. in *Advanced Visual Interfaces, AVI 2002*. 2002. Trento, Italy: ACM Press.
4. Pirolli, P., *Cognitive engineering models and cognitive architectures in human-computer interaction*, in *Handbook of applied cognition*, F.T. Durso, et al., Editors. 1999, John Wiley & Sons: West Sussex, England. p. 441-477.
5. Lynch, G., S. Palmiter, and C. Tilt. *The Max model: A standard web site user model*. in *Human Factors and the Web*. 1999.
6. Card, S.K., T.P. Moran, and A. Newell, *The psychology of human-computer interaction.* 1983, Hillsdale, New Jersey: Lawrence Erlbaum Associates.
7. Pirolli, P., *A web site user model should at least predict something about users.* internetworking, 2000. **3**.

8. Byrne, M.D., et al. *The tangled web we wove: A taskonomy of WWW use*. in *Human Factors in Computing Systems, CHI '99*. 1999. Pittsburgh, PA: ACM Press.

9. Blackmon, M.H., et al. *Cognitive Walkthrough for the Web*. in *Human Factors in Computing Systems, CHI 2002*. 2002. Minneapolis, MN: ACM Press.

10. Pirolli, P. *Computational models of information scent-following in a very large browsable text collection*. in *Conference on Human Factors in Computing Systems, CHI '97*. 1997. Atlanta, GA: Association for Computing Machinery.

11. Harman, D. *Overview of the first text retrieval conference*. in *16th Annuam International ACM/SIGIR Conference*. 1993. Pittsburgh, PA: ACM.

12. Card, S., et al. *Information scent as a driver of Web Behavior Graphs: Results of a protocol analysis method for web usability*. in *Human Factors in Computing Systems*. 2001. Seattle, WA.

13. Morrison, J.B., P. Pirolli, and S.K. Card. *A taxonomic analysis of what World Wide Web activities significantly impact people's decisions and actions*. in *Conference on Human Factors in Computing Systems, CHI '01*. 2001. Seattle, WA: ACM Press.

14. Ericsson, K.A. and H.A. Simon, *Protocol Analysis: Verbal reports as data*. 1984, Cambridge, MA: MIT Press.

15. Reeder, R.W., P. Pirolli, and S.K. Card. *Web-Eye Mapper and WebLogger: Tools for analyzing eye tracking data collected in web-use studies*. in *Human Factors in Computing Systems, CHI 01*. 2001. Seattle, WA.

16. McFadden, D., *Conditional logit analysis of qualitative choice behavior*, in *Frontiers of econometrics*, P. Zarembka, Editor. 1974, Academic Press: New York.

17. McFadden, D., *Modelling the choice of residential location*, in *Spatial interaction theory and planning models*, A. Karlqvist, et al., Editors. 1978, Harvard University Press: Cambridge, MA.

18. Bhavnani, S.K. *Domain-specific search strategies for the effective retrieval of healthcare and shopping information*. in *Conference on Human Factors nd Computing Sytems*. 2002. Minneapolis, MN: ACM Press.