

Running head: CoLiDeS and SNIF-ACT

## **CoLiDeS and SNIF-ACT: Complementary Models for Searching and Sensemaking on the Web**

Muneo Kitajima

National Institute of Advanced Industrial Science and Technology (AIST)  
Higashi Tsukuba Ibaraki 305-8566 Japan  
kitajima@ni.aist.go.jp

Peter G. Polson and Marilyn H. Blackmon  
University of Colorado at Boulder  
Boulder, Colorado 80309-0344 USA  
{peter.polson, marilyn.blackmon}@colorado.edu

### **Abstract**

CoLiDeS and SNIF-ACT are empirically-validated, complementary models. SNIF-ACT applies rational analyses to information foraging anywhere on the Web. CoLiDeS describes how people attend to and comprehend information patches on individual webpages. Integrating CoLiDeS and SNIF-ACT would better predict how people forage the Web for information to solve everyday ill-structured problems.

## CoLiDeS and SNIF-ACT: Complementary Models for Searching and Sensemaking on the Web

Information foraging theory depicts the human species as hungry for information, and as Pirolli (2005) has perceptively pointed out, navigating the Web has become a common way to find information needed to solve such ill-structured everyday problems as selecting treatment for a medical condition. Information foraging theorists have used ACT-R spreading activation models of information scent to generate reliable predications of how people navigate the Web by following an information scent trail. They have used mathematical models from rational analyses to calculate and compare utility values and accurately describe how people decide which particular information patch to graze in, when to select links to move to another webpage, when to back up to a previously visited information patch, and when to abandon a website and search for a new and hopefully better information patch (Pirolli & Card, 1999; Pirolli, 2005).

Although everyone has ill-structured everyday problems to solve, due to differences in background knowledge people vary enormously in their ability to comprehend the information available on the Web. Information foraging theory depicts people as hungry for information, but people in reality consume only information that they can comprehend. Information is useless to a person unless the person can comprehend the information. Due to differences in background knowledge people also vary in search strategies and attention management, ability to predict what links might be nested under superordinate categorical headings, and consequent ability to scan headings to identify what is in different patches. As a result of these differences in comprehension ability and attention management, people vary in both their ability to comprehend the information they find on the Web and their ability to find information by navigating the Web. Differences in background knowledge result from differences in culture, general reading knowledge, and amount of experience using Web browsers and computers, and researchers developing the CoLiDeS cognitive model and the Cognitive Walkthrough for the Web (CWW) have used the semantic spaces in Latent Semantic Analysis to generate reliable predictions for users that differ in general reading knowledge and culture (Blackmon, Mandalia, Kitajima, & Polson, 2007).

The SNIF-ACT model (Pirolli, 2005; Pirolli & Fu, 2003; Fu & Pirolli, in press), exemplifies the information foraging and rational analysis approach to predicting information search behavior anywhere on the Web. In contrast, the CoLiDeS (Kitajima, Blackmon, & Polson, 2000, 2005) model exemplifies the comprehension-based approach to predicting information foraging behavior at the microcosmic level of individual webpages, building bottom-up from the perspective of actions taken on an individual webpage. Whereas SNIF-ACT is founded on the ACT-R computational model, CoLiDeS is founded on Kintsch's (1998) Construction-Integration model of text comprehension, action planning, and problem solving – useful for understanding how people solve such ill-structured problems as comprehending and selecting treatment for a medical condition. The core argument of this paper will be that integrating these two complementary, empirically well-validated models of Web navigation – CoLiDeS and SNIF-ACT – would improve our ability to predict information search and sensemaking on the Web for the full gamut of human users of varying abilities.

### Information foraging anywhere on the Web

CoLiDeS and SNIF-ACT are complementary models, both starting with a user's goal to search for information. SNIF-ACT focuses on decisions to forage in a particular information patch, usually defined

as a complex website, or to leave the patch in search of patches with higher levels of information scent for the user's goal. As Figure 1 illustrates, an information patch can be defined at many different levels, from a particular website in the huge universe of websites on the Internet down to a collection of patches that compose a single webpage. SNIF-ACT computes the utility of staying within the current information patch compared to going back a page, clicking a link to go forward to a new page, or leaving the website. To date SNIF-ACT treats a webpage as a single information patch (Fu & Pirolli, in press), but there is no known barrier to extending SNIF-ACT to deal with a webpage as a collection of patches.

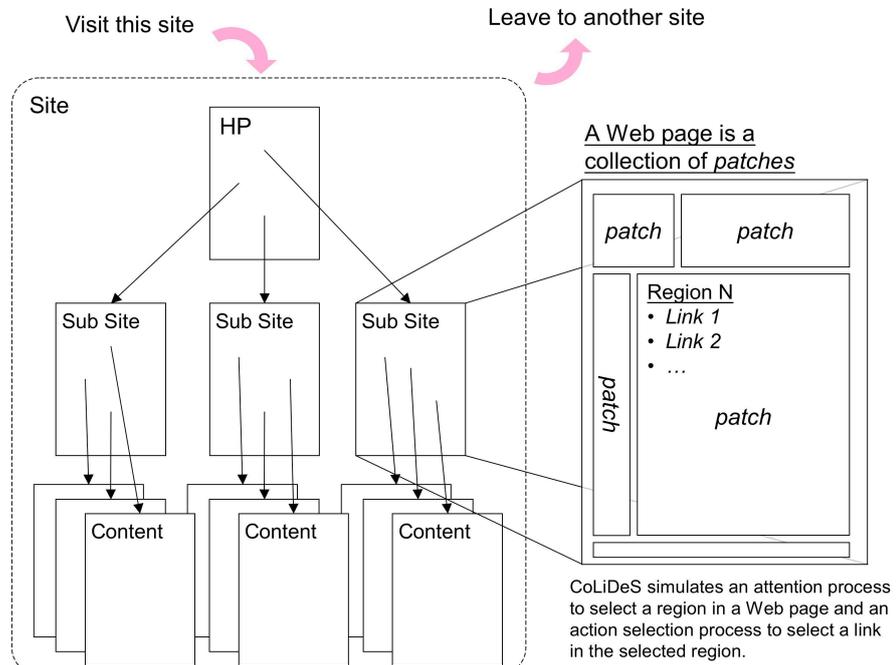


Figure 1. Information patches at all levels: individual websites within the universe of all websites on the Web, subsites, webpages, or patches within webpages

In contrast, CoLiDeS considers the current webpage as a collection of patches – called subregions in earlier publications (e.g., Kitajima, Blackmon, & Polson, 2005) – and uses information scent to select which particular patch to forage. Figure 2 illustrates a collection of patches on a single webpage. When either CoLiDeS or a human user is drawn to a patch with high information scent for the goal, the consequences are good if the patch actually contains a link that is on the solution path. In many cases, however, a human user is drawn to a patch with high information scent, where there are multiple high-scent links, none of which are on the solution path. This situation usually results in the user clicking many high-scent links that are not on the solution path (Blackmon et al, 2005, 2007). In these cases, information scent actively misleads the user, and the situation commonly occurs where items can be cross-classified but the Web designer makes the item accessible only by a link within one of categories.

Two other closely related problems can shackle persons who follow an information scent trail. One is the problem posed when a "correct patch" has relatively high scent but the "correct link" within that patch has very weak scent. Based on the mathematical models of rational utility of when to abandon a patch – and confirmed by empirical evidence (Blackmon, Kitajima, & Polson, 2005) – weak-scent correct links pose serious difficulties because people tend to abandon the patch without clicking the weak-scent correct link. The second closely related problem – discussed in the next section – is that clusters of links are often highly general categories that have some information scent for most goals but relatively low scent for any one particular goal. This dilemma calls attention to even deeper problems of how people select patches anywhere on the Web, because (a) website-level patches will have very low scent, except for webpages deep in the hierarchy listed in the results webpages of a search engine like Google, and (b) information found by search engines like Google is liable to be very unreliable despite high scent.

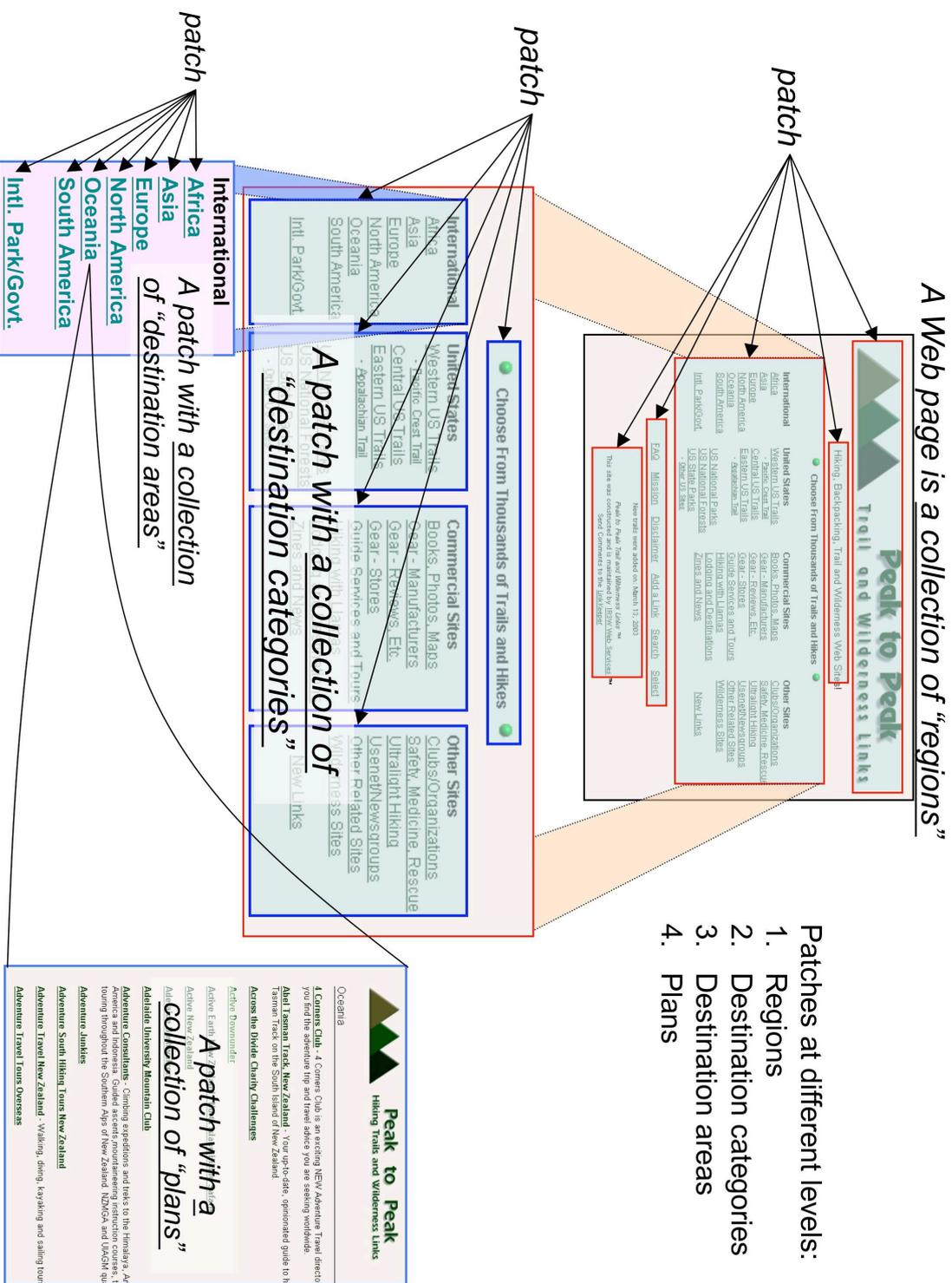


Figure 2. Patches within an individual webpage

## Sensemaking for information foraging on the Web

Information foraging theory draws from earlier models of foraging for food, and such models always take into account the nutritional content of the food – for example, calorie count, protein content, salt, minerals, and vitamins – and the tendency of organisms to avoid harmful constituents in the food source – for example, bacteria or toxins in the food or water that would cause the animal to become ill after eating the food. Sensemaking is a crucial element of information foraging, the analog of nutritional content of food and avoidance of constituents that would be harmful to the animal's health.

CoLiDeS is founded in the construction-integration architecture for text comprehension, action planning, and problem solving. Based on a theory of comprehension we can make three claims about sensemaking in Web navigation: (a) information discovered through information foraging is worthless to a person unless the person has the background knowledge required to comprehend it, (b) unreliable, untrue information is harmful and should be avoided, and (c) inability to adequately comprehend links, headings, and/or page layout conventions can seriously lower a person's success in finding the information needed or desired for solving an ill-structured everyday problem (see evidence on unfamiliar links reported in Blackmon, Kitajima & Polson, 2005). Figure 2 shows an example of an unfamiliar link, "Oceania," that is unfamiliar even for college-level readers. The Oceania link is liable to cause problems for a user searching for trails in New Zealand, because even college-level readers are unlikely to know that New Zealand can be considered part of Oceania.

**Comprehension of the information found.** In an extensive body of research, Kintsch has demonstrated the necessary role of background knowledge in constructing a situation model of the text. The situation model is required for text comprehension, for learning from text, for action planning and for problem solving (see review of this research in Kintsch, 1998). For example, in regard to finding information to solve the everyday ill-structured problem of finding information to select medical treatment, Patel and colleagues (e.g., Patel, Arocha, & Kushniruk, 2002) have documented patients' problems comprehending medical information about their condition, especially patients who have a narrative model of their disease and not a biomedical model like physicians and other medical professionals have.

**Reliability of the information found.** As Bhavnani et al. (2003) have argued, background knowledge is also crucial for determining the reliability of the information found, avoiding misleading/untrue information. Using searches on Google to find medical information is hazardous, because information available in the most reliable medical health websites (e.g., MayoClinic.com) is unlikely to appear on the first page of the results output by Google. Naïve searchers will be led to webpages whose information is unreliable, and only searchers with adequate background knowledge will be likely to restrict their search to the right information patches. Sophisticated searchers will either go directly to the most reliable websites for medical/health information, bypassing search engines entirely, or else they will be highly selective, clicking links to sites from respected institutions and accepting information as reliable only if it carries URAC- or HONcode-accreditation icons certifying that the developers of the website have agreed to follow a strict code of medical ethics in the information that they make available on the Web.

**Redefining information scent to include background knowledge factors in Web navigation by pure forward search.** Sufficient background knowledge is crucial not only for comprehending information found on the Web but also for successful navigation in search of information. Blackmon et al. (2002, 2003, 2005, 2007) have demonstrated the difficulties users have when they encounter a webpage whose links and headings are unfamiliar. In the latest iteration of CoLiDeS (Kitajima, Blackmon, & Polson, 2005), we combine five independent factors to determine the composite information scent between the user's goal and the screen object (i.e., heading or link text). CoLiDeS uses Latent Semantic Analysis (LSA) measures as engineering approximations for the first three of these five independent factors:

- The degree of semantic similarity between the user's goal and the heading/link text (LSA cosine)

- Whether there is an adequate level of relevant background knowledge for successfully elaborating the heading or link (minimum LSA term vector length in the selected semantic space estimates the amount of associated knowledge in the semantic space, e.g., amount of knowledge about diseases and conditions)
- Whether a word used in the heading or link text is a low-frequency term in the user's background knowledge (minimum LSA word frequency in the selected semantic space, a parameter that is especially crucial for very low frequency terms and zero-frequency terms typically ignored by LSA and by people)
- The frequency with which the user has encountered the screen object/widget or specific heading/link (screen elements on frequently navigated paths are more likely to be selected, e.g., a frequent user of websites with site navigation tab menus would have a propensity to navigate a website using the site tab menu, and, analogously, a person who had often used site search engines would be more apt to focus on the search window than someone who had previously located information primarily by browsing)
- Whether there is a literal matching, partial or complete, between the user's target goal and a screen object (e.g., looking for information about Type 2 Diabetes and seeing a link labeled "Type 2 Diabetes").

A full running simulation of CoLiDeS would integrate all the above five factors into a single activation value, i.e., a measure of the probability that the user will select a particular link or other screen object. It would also depend on constructing LSA semantic spaces that are psychologically valid representations of particular user populations, including differences in background knowledge and cultural meanings.

***Recently discovered issues in comprehension of navigation headings and links.*** New experiments in progress (Blackmon, unpublished data for Expt 060814) have recruited ethnic minority experimental participants, both children and parents, who have limited ability to read English and relatively low levels of educational attainment. The results from these experiments reveal serious difficulties that these participants have with abstract, very general superordinate categories. Ethnic minority participants had difficulty with selecting the correct top-level information patch in the website. Basic-level categories (e.g., Fish, Birds, Plants, Machines) were familiar to them and posed no problems, but they had difficulty with superordinate categories, such as "Social Science" or "Geography." Populations with college-level reading skills have no difficulty predicting what they will find when they select the link to such academic high-level categories as "Social Science" or "Geography," but preliminary analysis of this new dataset indicates that persons with lower levels of educational attainment (including children who read at the third-grade level) have little or no clue what subordinate links they can expect to find when they click the "Social Science" or "Geography" link. In the third-grade LSA semantic space, half of the links and headings in the experimental website, are *unfamiliar* as assessed by term vector length, and half of the words used in the headings and links are also low frequency words. The experimental website was an accurate simulation of the Encarta.msn.com online encyclopedia. It would be highly desirable if the individuals recruited for this experiment – with reading knowledge at the elementary school level (third- and sixth-grade LSA semantic space) – would be able to easily use online encyclopedias to help their children with school projects often assigned by school teachers for students in the upper elementary and secondary grades. Searching and sensemaking on the Web needs to be an opportunity that is equally available to all of our citizens, but we are currently far from that goal.

## Overview of talk

In sum, CoLiDeS and SNIF-ACT are complementary models, and integrating the two models would combine the rational analyses of information foraging on the Web with important issues of sensemaking in Web navigation that derive from the construction-integration theory of text comprehension, action planning, and problem solving. Our proposed talk would have two parts that focus on comprehension-based issues that derive from our research on the CoLiDeS cognitive model and the Cognitive Walkthrough for the Web (CWW) tool for evaluating webpage usability, one part on competing headings and weak-scent problems, relating to finding the right patch, and a second part on unfamiliar problems, focusing on understanding heading/link labels and content sufficiently to successfully navigate websites.

### 1.1 Part 1 of the talk: Scanning the webpage to identify the best information patch for deepening the search

Part one of our talk will address issues of scanning information patches on webpages and searching for goal-relevant information in hard-to-find patches. This includes what CWW originally called *competing headings problems* (where information scent lures people into the wrong information patch(es), *weak-scent correct links* (where people may find the right patch but abandon the patch before they click the correct link), and *weak-scent correct headings* (where the correct patch has little or no scent, providing no direction about where to focus attention). We could ultimately make contact with eye movement data to learn how people scan content and pay attention to information patches on a webpage. For the interim, redefining a webpage as a collection of patches spurs us to build into CoLiDeS the mathematical formulas from rational analyses to model (a) the decision to leave patch and move to another patch on the same webpage, (b) the decision to back up to previous webpage in the same website, (c) the decision to go to the home page of the website – or, in some cases, the main page of the subsite – and start top-down search over again, and (d) the decision to leave the website. In a nutshell, How can we design information patches within webpages so they have what Pirolli (2005, p. 347) calls "perfect information scent"?

#### 1.1.1 CoLiDeS and CWW on competing headings and weak-scent problems

For the task of finding an encyclopedia article about Dome of the Rock, a domed Muslim shrine in Jerusalem, the correct heading in the actual Encarta Encyclopedia was Art, Language and Literature, and the correct link was Architecture. Figure 3 shows the webpage with nine patches and its goal description for the Find Dome of the Rock task. CWW found a competing headings problem for this task, predicting a very strong pull towards several links nested under the heading Religion & Philosophy. Figure 4 shows

<p><b>Find encyclopedia article about Dome of the Rock</b>  <b>Dome of the Rock</b> (Arabic, Qubbat al-Sakhra), domed Muslim shrine in Jerusalem that stands on the traditional site of the Temple of Solomon (the first Jewish temple), the rock where, in the Biblical story of Abraham, Abraham had offered the sacrifice of his son Isaac to God. Figured to be the earliest surviving monument of Islamic architecture and probably modelled on the nearby Christian Church of the Holy Sepulchre, the Dome of the Rock was built not only to commemorate Muhammad's ascension to heaven, but also to rival the splendor of Christian and Jewish sanctuaries already in Jerusalem. The building is octagonal in plan, with a large golden dome on top (the original dome was metal covered with gold leaf, but a 1961 restoration replaced this with gold-colored anodized aluminum). . Centered under the dome, the Holy Rock itself may be seen, surrounded by an intricately carved wooden screen dating from 1199.</p>		
<p><b>Sports, Hobbies, &amp; Pets</b></p> <p><a href="#">Sports</a>  <a href="#">Sports Figures</a>  <a href="#">Games, Hobbies, &amp; Recreation</a>  <a href="#">Pets</a></p>	<p><b>Performing Arts</b></p> <p><a href="#">Theater</a>  <a href="#">Musicians &amp; Composers</a>  <a href="#">Cinema, Television, &amp; Broadcasting</a>  <a href="#">Music</a>  <a href="#">Dance</a>  <a href="#">Musical Instruments</a></p>	<p><b>Religion &amp; Philosophy</b></p> <p><a href="#">Theology &amp; Practices</a>  <a href="#">Mythology</a>  <a href="#">Religious Figures</a>  <a href="#">Philosophy</a>  <a href="#">Religions &amp; Religious Groups</a>  <a href="#">Scripture</a>  <a href="#">The Occult</a></p>
<p><b>Art, Language &amp; Literature</b></p> <p><a href="#">National &amp; Regional Literature</a>  <a href="#">Literature &amp; Writing</a>  <a href="#">Architecture</a>  <a href="#">Artists</a>  <a href="#">Language</a>  <a href="#">Writers &amp; Poets</a>  <a href="#">Decorative Arts</a>  <a href="#">Legends &amp; Folklore</a>  <a href="#">National &amp; Regional Art</a>  <a href="#">Painting, Drawing, &amp; Graphic Arts</a>  <a href="#">Sculpture</a>  <a href="#">Periods &amp; Styles</a>  <a href="#">Photography</a></p>	<p><b>Geography</b></p> <p><a href="#">World Cities, Towns, &amp; Villages</a>  <a href="#">Regions of the World</a>  <a href="#">Rivers, Lakes, &amp; Waterways</a>  <a href="#">Parks &amp; Monuments</a>  <a href="#">Countries</a>  <a href="#">Canadian Provinces &amp; Cities</a>  <a href="#">Islands</a>  <a href="#">Mountain Ranges, Peaks, &amp; Landforms</a>  <a href="#">U.S. Cities, Towns, &amp; Villages</a>  <a href="#">Maps &amp; Mapmaking</a>  <a href="#">Oceans &amp; Seas</a>  <a href="#">Exploration &amp; Explorers</a>  <a href="#">U.S. States, Territories, &amp; Regions</a></p>	<p><b>History</b></p> <p><a href="#">History of Asia &amp; Australasia</a>  <a href="#">People in European History</a>  <a href="#">People in United States History</a>  <a href="#">United States History</a>  <a href="#">African History</a>  <a href="#">World History &amp; Concepts</a>  <a href="#">Ancient History</a>  <a href="#">History of the Americas</a>  <a href="#">European History</a></p>
<p><b>Physical Science &amp; Technology</b></p> <p><a href="#">Construction &amp; Engineering</a>  <a href="#">Chemistry</a>  <a href="#">Earth Science</a>  <a href="#">Computer Science &amp; Electronics</a>  <a href="#">Machines &amp; Tools</a>  <a href="#">People in Physical Science</a>  <a href="#">Astronomy &amp; Space Science</a>  <a href="#">Paleontology</a>  <a href="#">Industr. Mining, &amp; Fuels</a>  <a href="#">Physics</a>  <a href="#">Transportation</a>  <a href="#">Communications</a>  <a href="#">Mathematics</a>  <a href="#">Military Technology</a>  <a href="#">Time, Weights, &amp; Measures</a></p>	<p><b>Life Science</b></p> <p><a href="#">Plants</a>  <a href="#">People in Life Science</a>  <a href="#">Medicine</a>  <a href="#">Invertebrate Animals</a>  <a href="#">Fish</a>  <a href="#">Algae &amp; Fungi</a>  <a href="#">Agriculture, Foodstuffs, &amp; Livestock</a>  <a href="#">Mammals</a>  <a href="#">Reptiles &amp; Amphibians</a>  <a href="#">Biological Principles &amp; Concepts</a>  <a href="#">Anatomy &amp; Physiology</a>  <a href="#">Environment</a>  <a href="#">Birds</a>  <a href="#">Viruses, Monerans, &amp; Protists</a></p>	<p><b>Social Science</b></p> <p><a href="#">Economics &amp; Business</a>  <a href="#">Organizations</a>  <a href="#">Institutions</a>  <a href="#">Political Science</a>  <a href="#">Psychology</a>  <a href="#">Law</a>  <a href="#">Education</a>  <a href="#">Anthropology</a>  <a href="#">Military</a>  <a href="#">Sociology &amp; Social Reform</a>  <a href="#">Calendar, Holidays, &amp; Festivals</a>  <a href="#">Archaeology</a></p>

Figure 3. Webpage for Dome of the Rock task, showing 9 information patches and goal statement.

a photo of Dome of the Rock. The goal-heading cosine for Religion & Philosophy was 0.59, compared to 0.05 for Art, Language and Literature. In addition to luring users down the garden path to a competing patch on the webpage (the Religion & Philosophy patch with its two *competing heading links*), the correct heading also has weak scent, minimizing the probability that users will ever focus attention on the Art, Language & Literature information patch before exiting the page or site.



Figure 4. Dome of the Rock, domed Muslim shrine in Jerusalem

As shown in Figure 5, the actual distribution of observed first clicks matches predictions. The red-shaded bars of Figure 5 show the one-level webpage design displayed in Figure 3. For this one-level webpage design, only 8 percent of the undergraduate experimental participants clicked a link nested under the correct heading, Art, Language & Literature. Most participants were lured down the garden path away from the correct heading/path, and 54% clicked a link nested under Religion & Philosophy, the predicted competing heading/patch, while another 26 percent clicked a link nested under Geography. Although Geography was not predicted to be a competing heading/patch, the correct heading had weak scent, and a weak-scent heading is expected to diffuse attention over other patches on the webpage. For the one-level webpage design shown in Figure 3, experimental participants were free to look at many patches before clicking a link in any particular patch. Because we do not have eye-movement data for the 76 undergraduates who did the Dome of the Rock task in the one-level design, we do not know where participants first focused their attention, only where they made their first actual click on a link.

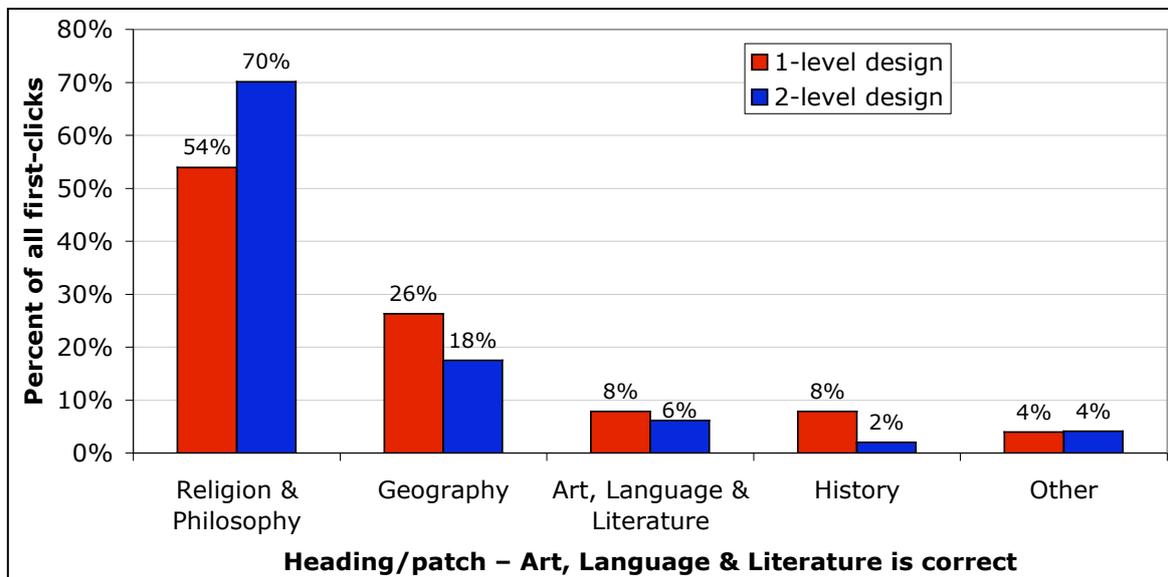


Figure 5. Observed clicks for undergraduate experimental participants doing the Dome of the Rock task on two different webpage designs: (a) 1-level design with all 9 headings/patches and all 93 links ( $n=76$  participants), and (b) 2-level design with 9 heading links visible on the top-level webpage and second-level webpages displaying all 9 headings plus all the links for one of the 9 headings ( $n=97$  participants)

In contrast, for the two-level webpage design (blue bars in Figure 5), participants were forced to click on one of the nine headings on the top-level webpage in order to see the links nested under that heading appear on the second-level webpage. Not surprisingly, for this two-level design, 70 percent of the 97 participants first clicked Religion & Philosophy, more than the 54 percent who clicked a link under Religion & Philosophy in the one-level design shown in Figure 3. Fewer clicked Geography (18 percent rather than 26 percent for the one-level design), and slightly fewer clicked the correct heading/patch, Art, Language & Literature (6 percent rather than 8 percent) or any other heading/patch. Two of the nine heading/patches — Performing Arts and Sports, Hobbies, & Pets — were never clicked by any of the 76 participants doing Dome of the Rock on the one-level design nor the 97 participants doing Dome of the Rock on the two-level design. The three science headings — Social Science, Life Science, and Physical Science & Technology — each had only a tiny percent of the clicks, never more than 3 percent.

The net effect was that 58% of the experimental participants time expired on the one-level version of the task, and 75% of the participants time expired on the two-level version of the task. Competing heading links and the weak-scent correct link combined to make this a very difficult task. This close-up view of a single task paves the way to understand the more abstract regression analysis of all 228 tasks in the data set we reported previously (Blackmon et al., 2005, 2007), and the expanded analysis of all 324 tasks in our complete data set of tasks simulating performance on the Encarta online encyclopedia, including 228 tasks analyzed by Blackmon et al. (2005, 2007), and 96 tasks reported earlier (Blackmon et al., 2002).

Number of tasks in dataset	Variable	Coefficient	Standard error	Standard coefficient	<i>t</i> -value	<i>p</i> -value
228 tasks (Blackmon et al., 2005)	Intercept	2.752	0.14	2.752	19.622	<.0001
	Competing heading links	0.797	0.063	0.641	12.567	<.0001
324 tasks: 228 + 96 other tasks	Intercept	2.484	0.111	2.484	22.363	<.0001
	Competing heading links	0.847	0.059	0.623	14.282	<.0001

For the 228-task dataset, the regression formula explains 41 percent of the data,  $F(1,226) = 157.93$ ,  $p < .0001$ , adjusted  $R^2 = .409$ , and the regression formula is  $Y = 2.752 + .797 * X$ . For the 324-task dataset, the regression formula explains 39 percent of the variance,  $F(1,322) = 203.973$ ,  $p < .0001$ , adjusted  $R^2 = .386$ , and the regression formula is  $Y = 2.484 + .847 * X$ .

It is crucial to compare this regression analysis for competing heading links with an analysis that regresses the observed clicks on the more standard variable of information scent, goal-link similarity, which CWW and CoLiDeS compute as the LSA cosine value for the goal statement text with the elaborated link text. Table 2 shows this alternate, far less successful attempt to explain the variance.

Number of tasks in dataset	Variable	Coefficient	Standard error	Standard coefficient	<i>t</i> -value	<i>p</i> -value
228 tasks (Blackmon et al., 2005)	Intercept	5.341	0.226	5.341	23.597	<.0001
	Cosine for goal with correct link	-5.699	0.632	-0.514	-9.018	<.0001
324 tasks: 228 + 96 other tasks	Intercept	4.772	0.204	4.772	23.37	<.0001
	Cosine for goal with correct link	-4.936	0.539	-0.455	-9.158	<.0001

For the 228-task dataset, the regression formula explains only 26 percent of the data,  $F(1,226) = 81.319$ ,  $p < .0001$ , adjusted  $R^2 = .261$ , and the regression formula is  $Y = 5.341 - 5.699 * X$ . For the 324-task dataset, the regression formula explains 20 percent of the variance,  $F(1,322) = 83.865$ ,  $p < .0001$ ,

adjusted  $R^2 = .204$ , and the regression formula is  $Y = 4.772 - 4.936 * X$ . The intercepts in Table 2 are much higher than the intercepts in Table 1, another inadequacy of the analysis using goal-link cosine as the independent variable rather than competing heading links. Adding a second variable to the regression, cosine for goal with correct heading, did not improve the analysis. The heading cosine variable is not statistically significant, the adjusted  $R^2$  was reduced, and the intercept coefficient remained very large.

On the other hand, adding weak-scent correct link to the regression analysis for competing heading links increases the percentage of variance, as displayed in Table 3. For the 228-task dataset, the regression formula explains 47 percent of the variance,  $F(1,226) = 99.886$ ,  $p < .0001$ , adjusted  $R^2 = .466$ . For the 324-task dataset, the regression formula explains 44 percent of the variance,  $F(1,322) = 125,983$ ,  $p < .0001$ , adjusted  $R^2 = .436$ .

Number of tasks in dataset	Variable	Coefficient	Standard error	Standard coefficient	t-value	p-value
228 tasks (Blackmon et al., 2005)	Intercept	2.539	0.14	2.539	18.14	<.0001
	Competing heading links	0.718	0.062	0.578	11.523	<.0001
	Weak-scent correct link	1.422	0.284	0.251	5.004	<.0001
324 tasks: 228 + 96 other tasks	Intercept	2.324	0.11	2.324	21.064	<.0001
	Competing heading links	0.755	0.059	0.555	12.749	<.0001
	Weak-scent correct link	1.455	0.267	0.238	5.456	<.0001

### 1.1.2 Implications for SNIF-ACT

The CoLiDeS and CWW analyses suggest that SNIF-ACT could be improved by not just processing links in sequential order on the webpage (Fu & Pirolli, in press) but instead paying serious attention to patch-level activation values (information scent) to explain scanning for and selecting some patches on webpages and ignoring other patches. Although the current version of SNIF-ACT did an excellent job of explaining the variance in a small dataset, a large percentage of webpages analyzed in that dataset were results pages from Google or other search engines, which are typically processed in sequential order. It is unlikely that SNIF-ACT could do a good job of explaining the variance in the 228-task or 324-task dataset analyzed here, where many usability problems are caused by the number of *competing heading links* and *weak-scent headings/links*.

## 1.2 Talk, Part 2: Insufficient background knowledge to comprehend

The second part of our talk will address insufficient background knowledge issues that CWW originally called *unfamiliar heading/link problems*. Unpacking these issues raises challenging problems. Is it possible to extend ACT-R to include a C-I theory of comprehension that accurately simulates the role of background knowledge and cultural differences in meaning? Is it at least feasible to input to SNIF-ACT numerical measures that match the more complex, five-component model of information scent that we have proposed for CoLiDeS?

### 1.2.1 CoLiDeS and CWW on unfamiliar problems

In our earliest experiments (Blackmon et al., 2002), we used webpages with randomly ordered collections of either 16 or 32 links drawn from the Encarta online encyclopedia (<http://encarta.msn.com>). These early experiments were launched to extend work by Larson and Czerwinski (1998) and used materials kindly shared with us by Larson and Czerwinski: webpages with 16 links arranged in a single column, and

webpages with 32 links arranged in two columns of 16 links each. Figure 3 shows that four of the tasks were more difficult, measured by mean solution time. All four of these tasks required selecting one of three links: (a) Anthropology (the tasks of finding Tlingit and Acculturation), (b) Paleontology (the task of finding Trilobite), or (c) Theology & Practices (the task of finding Canon Law). These three links are distinguished by low LSA term vector lengths, a measure of the quantity of closely related information stored in the selected LSA semantic space (the space representing background knowledge for a typical college freshman). Out of the 64 tasks, 12 required clicking one of the links with a low LSA term vector.

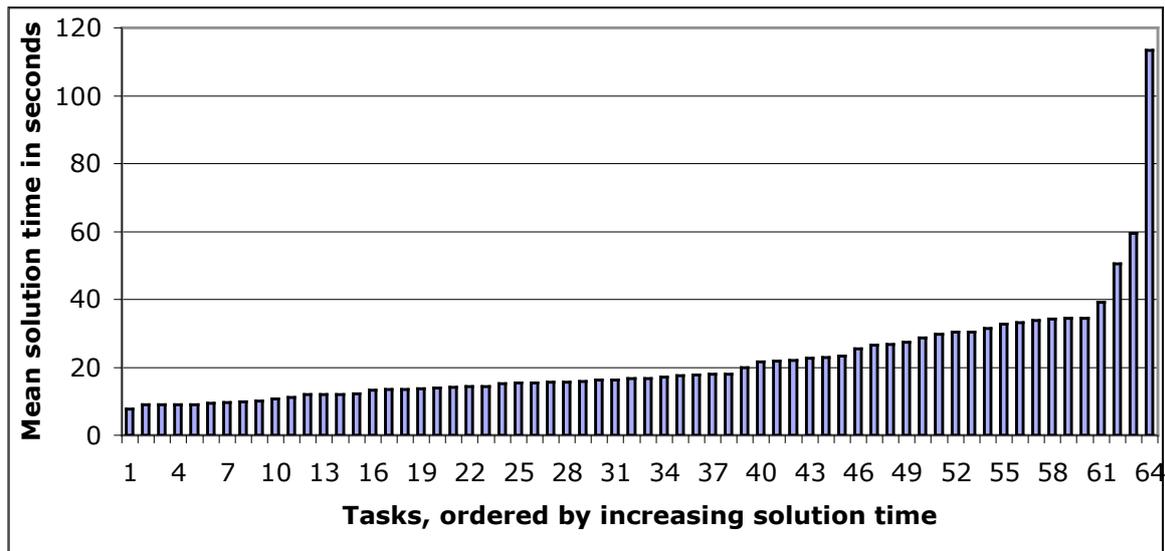


Figure 6. Mean solution time for 64 tasks done on webpages with no headings and either 16 or 32 links

The hardest of all the tasks was finding information about Tlingit, which required clicking the link "Anthropology" in the left-hand column of the 32-link webpage. Experimental participants were given the following description of Tlingit, which was an accurate summary of the actual Encarta Encyclopedia article about Tlingit:

Tlingit is the name of a group of Native American tribes of the Northwest Pacific Coast culture area and the Pacific coast of southeastern Alaska. The economy of the Tlingit is based mainly on fishing, and they are especially noted for their skill in woodcarving. In both appearance and social customs, they closely resemble the neighboring Haida. Today, the largest concentration of Tlingit is in Alaska, where many Tlingit work in the logging and fishing industries.

The mean solution time for finding Tlingit was 113 seconds, and 64 percent of the experimental participants time expired at 150 seconds without ever clicking the correct link to complete the task.

Figure 7 shows the mean solution time for all 64 tasks, separating the subset of tasks with unfamiliar correct link problems from the other 52 tasks, which were all non-problem tasks. There is a 2:1 ratio in mean solution times, and the difference is statistically significant,  $F(1,62) = 10.31$ ,  $p = .002$ .

Web navigation researchers are in consensus that users follow information scent when tracking down information, so we tested the set of 64 tasks to see if there was any difference in LSA cosine values between the unfamiliar set of 12 tasks and the familiar set of 52 tasks. Figure 8 shows the results of this test, finding virtually identical means for the two subsets and no statistically significant differences.

Extending the analysis to the full sample of 324 tasks, there is also a significant difference in mean number of clicks between unfamiliar problem tasks and other tasks. The mean total clicks for familiar

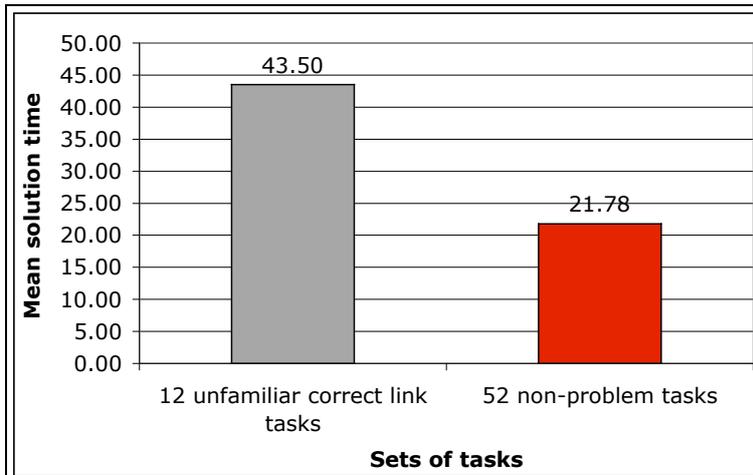


Figure 7. Mean solution time for 12 tasks with an unfamiliar correct link problem versus 52 non-problem tasks

tasks ( $n=268$ ) = 2.9, and the mean total clicks for unfamiliar tasks ( $n=56$ ) = 4.9,  $F(1,322) = 39.205$ ,  $p < .0001$ .

The next step is to add the unfamiliar correct link to the multiple regression analysis, as shown in Table 4. For both the 228-task dataset and the full 324-task dataset three independent variables explain over 50 percent of the variance, and the coefficients are highly similar for both datasets and for the smaller datasets reported in Blackmon et al. (2005). For the 228-task dataset the tasks are relatively homogenous, and the

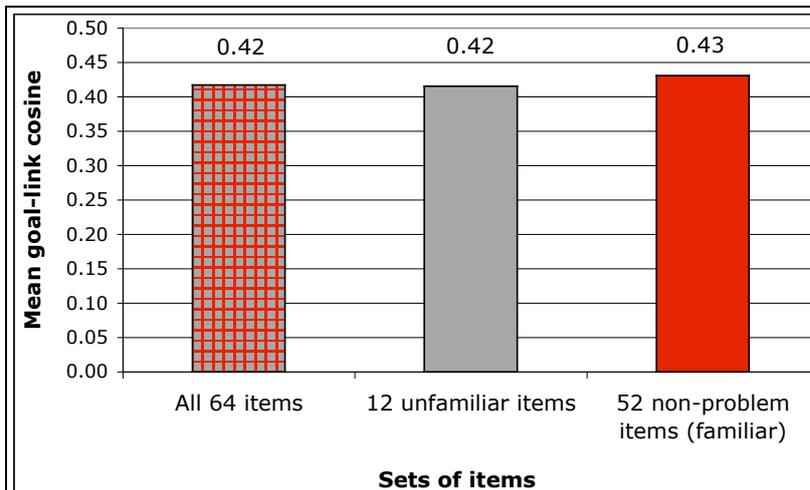


Figure 8. Mean LSA goal-link cosine for 12 tasks with an unfamiliar correct link problem vs. 52 non-problem tasks vs. the full set of 64 tasks.

three variables explain 54 percent of the variance,  $F(3,224) = 89.961$ ,  $p < .0001$ ,  $R^2 = .54$ . For the 324-set dataset the three variables explain 51% of the variance,  $F(3, 320) = 110.903$ ,  $p < .0001$ ,  $R^2 = .51$ .

Published data by Blackmon et al. (2002, 2003, 2005, 2007) have been limited to college-level reading knowledge and undergraduate experimental participants who are generally quite computer literate. Therefore, even college-level users are seriously hampered by unfamiliar problems, and we

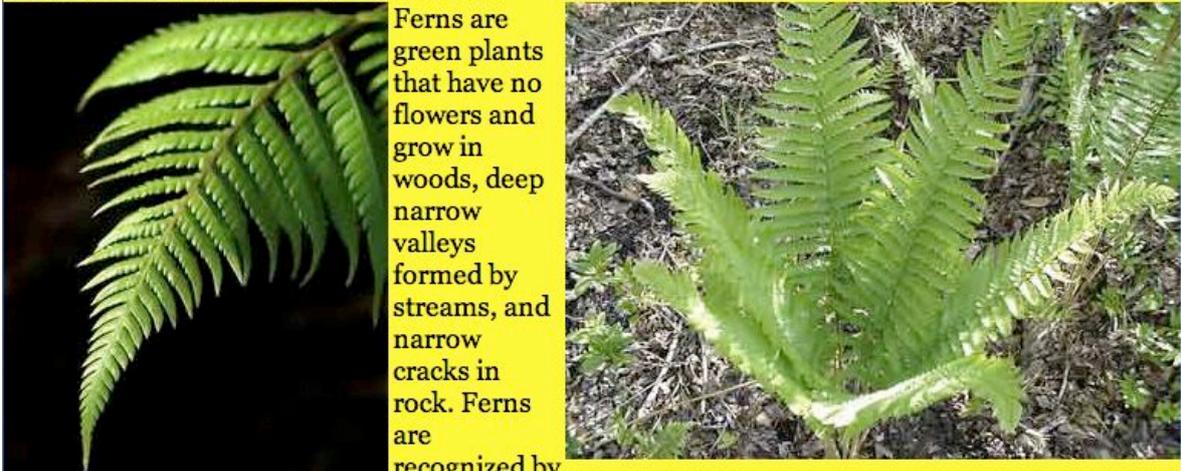
Number of tasks in dataset	Variable	Coefficient	Standard error	Standard coefficient	t-value	p-value
228 tasks (Blackmon et al., 2005)	Intercept	2.309	0.135	2.309	17.085	<.0001
	Competing heading links	0.661	0.059	0.532	11.298	<.0001
	Weak-scent correct link	1.746	0.285	0.28	6.133	<.0001
	Unfamiliar correct link	1.45	0.264	0.256	5.5	<.0001
324 tasks: 228 + 96 other tasks	Intercept	2.079	0.11	2.079	18.973	<.0001
	Competing heading links	0.707	0.056	0.52	12.625	<.0001
	Weak-scent correct link	1.523	0.25	0.249	6.09	<.0001
	Unfamiliar correct link	1.597	0.236	0.267	6.758	<.0001

can expect far more serious difficulties to be incurred by users with much lower levels of background knowledge, and lower levels of computer literacy.

Our most recent experiments concern ethnic minority families, both adults and children, who have limited computer literacy and limited comprehension of English at levels as low as third-grade general reading knowledge. Based on CWW analyses using the third-grade semantic space, we anticipate that about half of the 93 links on the Encarta encyclopedia are unfamiliar links, and about half of the words used in the links are low frequency words. Nevertheless, the tasks themselves required searching for information about familiar items, and the goal descriptions provided pictures of the items they should search for, such as photographs of a fish species called Marlin, or photographs of ferns for the Find Fern task shown in Figure 9. It is reasonable to suppose that low literacy users might be able to ignore the unfamiliar links and still click familiar ones. For example, for the Find Fern task they might first go to the Life Science information patch and then scan for and click the link Plants in that patch. Nevertheless, even seemingly simple tasks like this proved surprisingly difficult for this population.

College-level participants clearly understand what they will get if they click Geography or Social Science or Religion & Philosophy or Performing Arts, but these lower literacy users often select such headings as Geography or Social Science using a trial-and-error strategy. Indeed, CoLiDeS and CWW predict this behavior for this population, because for the third-grade semantic space these four headings are all *unfamiliar headings*. The low term vectors for these four headings indicate that for individuals with third-grade general reading knowledge such headings as Geography or Social Science are vague categories and are not very meaningful. These four unfamiliar headings are analogous to the vague concepts that college-level readers have for the *unfamiliar links* Paleontology or Anthropology, the unfamiliar link problems that caused Tlingit and Acculturation to be such difficult tasks for college-level participants in our early experiments. In addition, although the ethnic minority participants in our latest experiment are familiar with the word "science," they appeared to be puzzled and confused by what might be the difference

**Homework: Find article about Fern**



Ferns are green plants that have no flowers and grow in woods, deep narrow valleys formed by streams, and narrow cracks in rock. Ferns are recognized by the feather-like shape of their leaves. There are about 12,000 species of ferns throughout the world. Ferns grow mostly in tropical areas. The Arctic and Antarctic regions of the Earth have only a small number of ferns. Many ferns grow on trunks and branches of trees. Ferns range from tiny plants to tall tree ferns that measure 30 to 80 feet high. Life cycle Ferns were present on Earth before plants with flowers. Ferns reproduce with spores and produce both male sperm and female eggs needed to produce new plants.

Figure 9. Goal description for the Find Fern task uses photographs and uses only word familiar to someone with a third-grade level of general reading knowledge.

between Life Science, Social Science, and Physical Science & Technology. The term vectors for these terms are marginally sufficient, but they still appeared to struggle with these headings, too. Analyses of this new set of experiments is in progress and will be reported in the HCIC talk.

### 1.2.2 Implications of unfamiliar problems for SNIF-ACT

At the present time SNIF-ACT ignores unfamiliar problems, relying only on semantic similarity measures, and we argue that the next version of SNIF-ACT should be revised to take a comprehension-based approach that takes background knowledge of users into account. When users do not have sufficient background knowledge to comprehend links or headings, they are liable to not be able to find the information they are searching for on the Web. Concepts that seem so obvious to a college-educated web designer are vague and not meaningful to someone with a third-grade reading knowledge. Unassisted by CoLiDeS and CWW, we could never have predicted the behaviors seen in the latest experiment, and SNIF-ACT can be set up to make the same predictions.

## 2 Conclusions

CoLiDeS and SNIF-ACT are like two trains approaching each other from opposite directions. CoLiDeS, ground in Kintsch's (1998) construction-integration theory of comprehension has focused on individual webpages. SNIF-ACT has worked at a higher level, paying little attention to differences at the microscopic level of individual webpages but providing good explanations of navigation from one webpage to another and one website to another. The two models complement each other and each can benefit from incorporating the specialized strengths of the other.

Scanning a webpage to grasp its structure requires the ability to segment the webpage into information patches, as illustrated in Figures 10a and 10b. This is one of the specialized strengths of CoLiDeS that is highly compatible with Information Foraging theory.

Unfamiliar problems are widely recognized, particularly by educators, but what is unique here is the demonstration that semantic similarity — what most theorists call information scent — is largely independent of unfamiliarity. Unfamiliar links may have high LSA cosines, i.e., high measures of semantic similarity, but the semantic similarity is not perceptible to someone who does not have sufficient background knowledge to comprehend it.

Weak scent has consistently proved to be a statistically significant determinant of performance in CWW research, but it may receive its best interpretation in Information Foraging theory, where users are predicted to tend to leave the information patch before they exhaustively search all links. According to Information Foraging theory, users are not expected to be particularly bothered by having multiple attractive links within the correct information patch, but they will be derailed if the correct link in that patch has very weak scent.

Finally, we have looked at the issue of sequential processing, trying to find evidence for it in various subsets of the data. Although it seems plausible to assume that users would scan a column of links from top to bottom, there is no evidence that the position of the link in the column affects the number of clicks required to locate the link. There is some evidence to indicate that it takes more clicks to select a link located in the right-hand column of 16 links on the webpages with 32 links compared to the left-hand column of 16 links or the at-link webpages. Such processing orders may, nevertheless, prove important in eye-movement data.

More generally, we will argue that comprehension and sensemaking are crucial and have to date not been given sufficient emphasis in HCI theories.



Figure 10a and 10b, showing patch structure of New York Times webpage

## References

- Bhavnani, S.K., Bichakjian, C.K., Johnson, T.M., Little, R.J., Peck, F.A., Schwartz, J.L., & Strecher, V.J. (2003) Strategy Hubs: Next-generation domain portals with search procedures. *Proceedings of CHI 2003* (pp. 393–400), New York: ACM Press.
- Blackmon, M. H., Mandalia, D. R., Kitajima, M., & Polson, P. G. (to appear, 2007). Automating Usability Evaluation: Cognitive Walkthrough for the Web Puts LSA to Work on Real-World HCI Design Problems. In T. K. Landauer, D. S. McNamara, S. J. Dennis, and W. Kintsch (Eds.), *The Handbook of Latent Semantic Analysis: A Road to Meaning* (chapter 18). Mahwah, NJ: Erlbaum.
- Blackmon, M. H., Kitajima, M., & Polson, P.G. (2003) Repairing usability problems identified by the Cognitive Walkthrough for the Web. *Proceedings of CHI 2003* (pp. 497–504). New York: ACM Press.
- Blackmon, M. H., Kitajima, M., & Polson, P. G. (2005). Tool for accurately predicting website navigation problems, non-problems, problem severity, and effectiveness of repairs. *Proceedings of CHI 2005* (pp. 31–41). New York: ACM Press.
- Blackmon, M. H., Polson, P. G., Kitajima, M., & Lewis, C. (2002). Cognitive Walkthrough for the Web. *Proceedings of CHI 2002* (pp. 463–470). New York: ACM Press.
- Fu, W.-T., & Pirolli, P. (in press). SNIF-ACT: A cognitive model of user navigation on the World Wide Web. Accepted for publication in *Human-Computer Interaction*.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, U.K. & New York: Cambridge University Press.
- Kitajima, M., Blackmon, M. H., & Polson, P. G. (2000). A Comprehension-based model of Web navigation and its application to Web usability analysis. *Proceedings of HCI 2000* (pp. 357–373). Springer-Verlag.
- Kitajima, M., Blackmon, M. H., & Polson, P. G. (2005). Cognitive architecture for website design and usability evaluation: Comprehension and information scent in performing by exploration. *Proceedings of HCI-International 2005*, Vol. 4: *Theories, Models, and Processes in HCI* (paper 1866).
- Larson, K. & Czerwinski, M. (1998). Web page design: Implications of memory, structure and scent for information retrieval. In *Proceedings of CHI'98* (pp. 25–32). New York, NY: ACM Press.
- Patel, V. L., Arocha, J. F., & Kushniruk, A. W. (2002). Patients' and physicians' understanding of health and biomedical concepts: Relationship to the design of EMR systems. *Journal of Biomedical Informatics*, 35, 8–16
- Pirolli, P. (2005). Rational analyses of information foraging on the Web. *Cognitive Science*, 29, 343–373.
- Pirolli, P., & Card, S. K. (1999). Information foraging. *Psychological Review*, 106, 643–675.
- Pirolli, P. L., & Fu, W.-T. (2003) SNIF-ACT: a model of information foraging on the World Wide Web (9th International Conference on User Modeling). *Lecture Notes in Artificial Intelligence*, 2702, 45–54.