

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Integrating graphical information into cognitive modeling of web navigation

Permalink

<https://escholarship.org/uc/item/72k5t9rh>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 31(31)

ISSN

1069-7977

Authors

Indurkha, Bipin
Karanam, Saraschandra
Puerta Melguizo, Mari Carmen
et al.

Publication Date

2009

Peer reviewed

Integrating graphical information into cognitive modeling of web navigation

Saraschandra Karanam (saraschandra@research.iiit.ac.in)

Cog Sci Lab, International Institute of Information Technology, Gachibowli,
Hyderabad, India.

Herre van Oostendorp (herre@cs.uu.nl)

Center for Content and Knowledge Engineering, Institute of Information and Computing Sciences,
Utrecht University, Utrecht, Netherlands.

Mari Carmen Puerta Melguizo (puerta@cs.uu.nl)

Center for Content and Knowledge Engineering, Institute of Information and Computing Sciences,
Utrecht University, Utrecht, Netherlands.

Bipin Indurkha (bipin@iiit.ac.in)

Cog Sci Lab, International Institute of Information Technology, Gachibowli,
Hyderabad, India

Abstract

Cognitive models of web navigation like CoLiDeS, use only textual information from hyperlinks to compute information scent and ignore so far the impact of visual and graphical widgets. We conducted an experiment to study the extent to which textual and especially graphical information, plays a role in identifying web page widgets. Four different versions of a webpage were created by systematically varying text and graphics. In general, task completion times and number of clicks were significantly less in the presence of graphics than in their absence. This was particularly the case when there was no textual information available. We conclude that for identifying graphical widgets, text and graphics interact and complement each other and it is important for a cognitive model on web navigation to include information from graphics. In this direction we propose a method to integrate information extracted from pictures into CoLiDeS and we demonstrate its usefulness with a simulation done on a mock web site.

Keywords: Web navigation; web usability; cognitive model; information scent; semantics; graphics

Introduction

Navigating the Internet and searching for information requires the processing of (at least) visual and textual information (Paivio, 1986). Cognitive modeling of web navigation behaviour has been extensively studied and a number of models have been developed. However, these models do not pay much attention to the role of visual or graphical information. Kitajima, Blackmon, and Polson (2000) developed a theoretical model of web-navigation called *Comprehension-based Linked model of Deliberate Search* (CoLiDeS). This model assumes that comprehension of texts and images is the key to web navigation. Comprehension processes build, elaborate and compare the mental representations of screen objects to determine which hyperlink or image to select and click. This comparison involves computing semantic similarity between the search goal and the screen objects (Pirolli & Card, 1999).

To predict navigational choices, Pirolli and Fu (2003) developed an architecture called *Scent-based Navigation and Information Foraging in ACT* (SNIF-ACT). This architecture considers the whole of world-wide web as a semantic space, and predicts navigational choices — such as which link to click, where to go next, when to leave the website, and so on — based on a parameter called *information scent*, which is calculated as the mutual relevance between the user goals and the link texts based on word occurrences and co-occurrences on the internet.

Miller and Remington (2004) proposed a *Method for Evaluating Site Architectures* (MESA), which gives a model for explaining user backtracking behaviour. It models various navigation styles and strategies to recover from selecting misleading links, and gives an account of the effectiveness of selecting various links given their relevance to the search goal.

Finally, Juvina and Van Oostendorp (2007) and Van Oostendorp and Juvina (2008) developed CoLiDeS+, which extends CoLiDeS by including contextual information such as information from the selected links on previously visited web pages, namely the navigation path. CoLiDeS+ defines *path adequacy* as the semantic similarity between the navigation path and the goal. When the incoming information from links on the current page increases path adequacy, it is considered for selection, otherwise alternate paths are chosen.

All these models are based on the concept of semantic similarity or *information scent* since it has been verified that this drives the user's search behaviours and navigation patterns (Chi *et al.*, 2000, 2001, Pirolli & Card, 1999). CoLiDeS and CoLiDeS+ use a mathematical technique called *Latent Semantic Analysis* (LSA) developed by Landauer *et al.* (1998). LSA estimates semantic relatedness of texts, based on a statistical analysis of large corpus. These similarity measures between user-goals and hyperlinks on a webpage, as given by LSA, are used to predict the likelihood of a user selecting each of those hyperlinks.

All these models, however, seem to ignore the semantic information gained from the visual/graphical modality. Although the models acknowledge the importance of comprehension of images in web navigation, CoLiDeS computes the influence of semantics derived from only text. Research on visual search and visual saliency highlights the fact that our visual system is not only adept at perceiving salient objects (Desimone & Duncan, 1995; Itti & Koch, 2001) but also quicker (Paivio, 1986; Mayer & Moreno, 2003). Although it seems worthwhile to study the impact of semantics derived from graphics in the domain of web navigation, there is almost no existing research on it. In this paper, we explore the role of graphical and visual information, in contrast with textual information, in locating web page widgets.

Hinesley (2005) examined the impact of graphics in locating widgets on a webpage by taking original and greeked web pages (pages in which any textual information has been substituted by sequences of X's, as shown in the Figure 1). She found that in the absence of textual information (greeked pages), widgets with mostly graphical information (e.g., a conventional search box, advertisement) were found more quickly than widgets that were purely textual (e.g., Contact Link, Privacy Statement). In a second experiment Hinesley and Blackmon (2008) systematically varied two variables: graphics (Graphics vs. No Graphics) and location expectations (Location vs. No Location) on greeked pages. The loss of graphical information was found to have twice as large an impact as the violation of location expectations. This research claims that graphics, and not text, seems to be the key to how users find and recognize popular conventional graphical widgets like the search engine.

While text was manipulated in the presence of graphics in one experiment, graphics was manipulated in greeked text in another. The role of text in finding graphical widgets was not addressed in Hinesley's work. This forms the main focus of our first study. We hypothesize that the difference between graphics and no-graphics conditions is smaller in the presence of textual information.

In the second study we present the results of our preliminary simulation of CoLiDeS including pictures. We hypothesize that when including semantic information from graphics into the model, the right navigation path can be found more frequently and more accurately than when it is not included.

Study 1

Participants

Forty students, from Utrecht University, including post-graduate and research scholars, participated in the experiment.

Design

We used a 2 (text vs. no text) X 2 (graphics vs. no graphics) X 12 Widget repeated measures design. The dependent

variables were the mean task completion time and the number of clicks a user takes to locate the correct widget.

Material and Apparatus

Four versions of each webpage were created using a 2 (text vs. no text) X 2 (graphics vs. no graphics) manipulation – (1) Normal Version with both text and graphics intact (T+ G+), (2) No Text Version with text removed (greeked) and graphics intact (T- G+), (3) No Graphics Version with graphics removed and text intact (T+ G-), and (4) No Graphics and No Text Version with both text and graphics removed (T- G-). Other variables such as size and location were maintained constant across all versions of 8 web pages. For the no text versions, all text was replaced with character “X” in the same font, spacing and style. For the no graphics versions, all images were removed leaving the text in Arial typeface with 8-point size in the original positions. Fig1 shows a small portion of a webpage in all 4 versions.



Figure 1: Four versions of a webpage

All widgets were divided into purely textual widgets (Contact Us, Privacy Statement, Newsletter Sign Up, Left Navigation, Top Navigation and Register/Login) and graphical widgets (Accreditation, Logo, Search Engine, Advertisement, Print Icon and Drop-Down Menu). Some widgets were mostly graphical in the sense that they had some text in fancier format (the text “Go” for a search engine). These were also classified under graphical widgets. An image map was created for each web page, based on functional web units, assigning a clickable region to each object so that when a user hovered over any widget, the mouse pointer changed to a hand, indicating that the user could click. Participants were tested using a Dell 17” flat-panel displays running Windows XP with SP2. Browser used was Internet Explorer 7.0.

Procedure

Participants performed a total of 56 trials on eight different web pages. Their task was to locate widgets and click on them. Each participant was asked to locate six widgets (three textual and three graphical) on one version of each of the eight websites. Participants first saw the task description on the screen and then started performing the tasks. A description of the task was always present on top of the page. Care was taken to ensure that each participant saw only one version of any webpage.

Results

Task Completion Time A 2X2 within subjects ANOVA was conducted with text and graphics as independent variables and mean task completion time as dependent variable. Results show a main effect of graphics ($F(1,39) = 28.17, p < .001$). The task completion times are significantly less in the presence of graphics when compared to the times with no graphics. The main effect of text is not significant ($p > .05$). The interaction of text and graphics is significant ($F(1,39) = 5.83, p < .05$). Figure 2 shows the interaction between text and graphics.

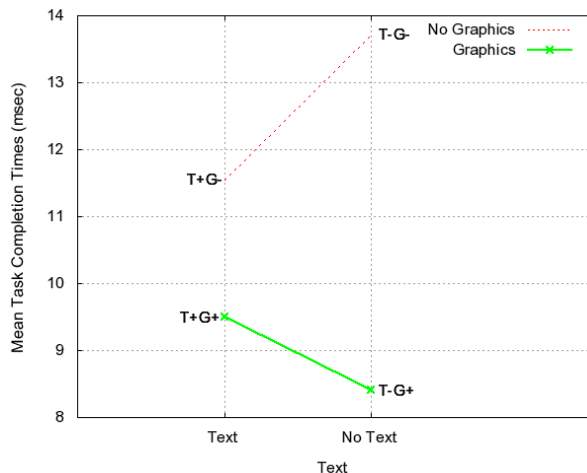


Figure 2: Mean Task Completion Times in relation to text and graphics

T-tests reveal significant differences between all pairs except (T+ G+)-(T- G+). Moreover, the difference between (T-G+)-(T-G-) is much greater than the difference between (T+G+)-(T+G-). These values support our hypothesis that that the difference between graphics and no-graphics conditions is smaller in the presence of textual information. Also, under the no-graphics condition, participants took significantly less time in locating widgets in the presence of text than in its absence. Semantic information from text helped users in locating widgets when no graphics are presented.

Number of Clicks A similar 2X2 within subjects ANOVA was conducted with text and graphics as independent variables and the number of clicks taken to find the right

widget as dependent variable. The main effect of text is significant ($F(1,39) = 49.70, p < .001$). The number of clicks is significantly greater in the absence of text. Also, the main effect of graphics is significant ($F(1,39) = 37.06, p < .001$). The number of clicks in the absence of graphics is significantly greater than in the presence of graphics. Finally, the interaction of text and graphics is also significant ($F(1,39) = 35.6, p < .001$). Figure 3 depicts this relationship for average number of clicks.

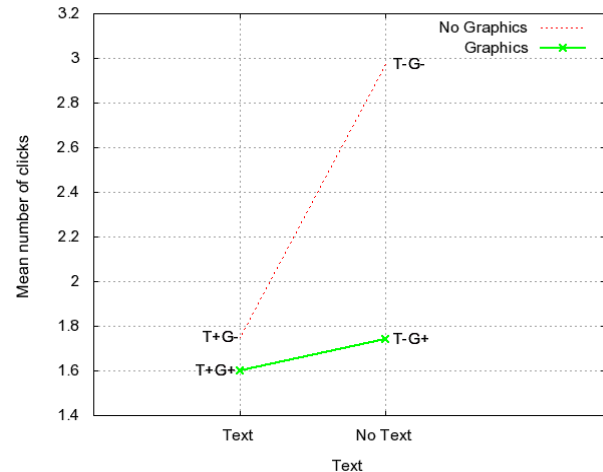


Figure 3: Mean number of clicks in relation to text and graphics

T-tests reveal two significant pairs (T- G+)-(T- G-) and (T+ G-)-(T- G-). These effects emphasize the importance of semantic information derived from text in locating widgets since the comparison between the conditions (T+ G+) and (T+ G-) is not significant while (T- G+) compared with (T- G-) is highly significant. Furthermore, the effect of graphics is significant only in the absence of text.

Nature of Task A 2X2X2 within subjects ANOVA was performed with text, graphics and task type (localising textual vs. graphical widgets) as independent variables and mean task completion times and mean number of clicks as dependent variables.

Both for task completion times and number of clicks, graphical tasks followed the same pattern as in Figure 2 and 3. This implies that for graphical tasks, graphics are very important. For textual tasks, presence or absence of either text or graphics does not play any significant role.

Discussion

The results of this experiment show that both text and graphics play an equally important role in locating webpage widgets. Consequently, in contrast to Hinesley and Blackmon (2008), it is not graphics alone, but both sources that are important for identifying objects. They both interact in the sense that text assumes a greater role in the absence of graphics, and graphics assumes a greater role in the absence

of text. Removing both text and graphics results in the worst performance results.

The implication of this study for cognitive models of web navigation behaviour is that while text is important, graphical information is equally important. This effect is especially prominent for conventional graphical widgets. Consequently, we claim that models of web navigation would predict navigation behaviour more accurately if the information from both text and graphics is taken into consideration in computing *information scent*. In the second study we present a way to do it and evaluate our results.

CoLiDeS + Pic

We propose an extension of CoLiDeS to include the semantic information coming from pictures in a webpage in addition to the information scent computed from hyperlinks. We call it CoLiDeS + Pic as it shares all the main assumptions of the original cognitive model with only a few changes at the level of computation of similarity.

CoLiDeS + Pic is based on the hypothesis that the presence of pictures that are semantically close to the current context would aid the user in selecting the correct link predicted by CoLiDeS. Context is provided by the user goal, visible hyperlinks and the picture.

Cognitive grounds of CoLiDeS + Pic

Though CoLiDeS assumes that comprehension of text and images is key to web navigation, it never modeled images. CoLiDeS + Pic takes the first steps in that direction.

CoLiDeS considers that all different objects in the website are related to each other and to the user goal by three measures of relevance – similarity, frequency and literal matching. The similarity measures used by CoLiDeS compute the similarity between the user goal and each of the objects separately and consider only textual objects. CoLiDeS + Pic fills this gap to some extent by including semantic information coming from pictures, thereby making the model more consistent with the theoretical assumptions.

According to CoLiDeS, selecting an object on a webpage is an outcome of four cognitive processes – *parsing* all the objects contained in the webpage into 5-10 top level schematic objects (e.g. Window Controls, Left Navigation Column, Top Navigation Column), *focusing* on one of the top level schematic objects, *comprehending and elaborating* the objects within the focused area and *selecting* one of them as a target for the next action. CoLiDeS + Pic in a way combines the two phases of elaboration and focusing on. It is during the parsing phase that the user glances at all the objects. After elaborating each of the objects, the user decides to focus on one of them based on three measures of relevance. By including elaborations done for pictures along with elaborations done for hyperlinks, CoLiDeS + Pic is using more information from the elaboration phase to decide which final object to select. Thus, CoLiDeS + Pic provides a more accurate cognitive model.

First we describe the method we use to extract semantic information from pictures. This extraction method is to our knowledge not used before, and it is in our opinion interesting to examine whether it leads to useful results. Next we describe a simulation study, which is based on CWW - Cognitive Walkthrough for the Web (Blackmon *et al.*, 2002, 2005). CWW is a usability inspection method that identifies problems in web-page design. Our method adds certain additional steps to CWW to incorporate semantic information from pictures.

Steps involved in extracting semantics from pictures

According to the Construction Integration Theory of Text-Comprehension (Kintsch, 1998), as the reader proceeds through a piece of text, he/she constructs a mental representation of the incoming piece of text by using his background and domain knowledge. Only a small portion of human memory is active at any point of time. The incoming text element, the previously read text, user goal and user background knowledge determine in that order of priority which concepts are active at any given point of time. Similarly, when a user looks at a picture, many concepts are activated, influencing the interpretation of other information in that context, for instance, the hyperlinks. So we assume that the activated features co-determine the link that will be chosen. Thus, the first step is to extract the semantic features that a picture activates in user's memory.

User goals in our case correspond to the questions the user needs to answer (e.g. "Name three layers of tissue that form the heart wall?"). For each user goal, five pictures varying in degree of relevance to the context of those pages are collected. Participants are then asked to write down five semantic features based on the concepts that come to their mind looking at the picture in context (See Figure 4 for an example).

By combining the features from participants, we obtain a collection of semantic features for each picture. By selecting common features given by two or more users, we generate a frequency distribution of these re-occurring features. A feature is deemed to be representing the picture if it is among the top five in the frequency distribution and is mentioned by at least 50% of all the users. The picture is now represented by these top five most frequently mentioned features.



Figure 4: Snapshot of Webpage with High relevant picture used for Feature Extraction

High and Low Relevant Pictures For each page, LSA is used to compute the semantic similarity between the goal and five sets of concatenated features representing one of the 5 pictures. The picture corresponding to the highest cosine value is taken to be the high relevant picture and the picture corresponding to the lowest cosine value as the low relevant picture.

Integrating semantic information from pictures into CoLiDeS CWW is a usability inspection method that attempts to account for the four phases of CoLiDeS (parsing, comprehension and elaboration, focusing on and selecting). CWW assumes that a user selectively attends to that sub-region of a page whose description is most similar to her goal. It uses LSA to compute this similarity. The main steps involved in CWW (Blackmon *et al.*, 2002) are: for each goal selecting a semantic space that best fits the user profile, compiling a set of realistic user goals, elaborating the goal and the links, and finally estimating semantic similarity of goal and links. In our approach, these four steps remain the same. As a fifth step, we collect semantic features from pictures relevant to each goal, elaborate the features obtained from pictures along with elaboration of goal and the links. The elaborated features are appended to the elaborated links. Final step is to compute semantic similarity between the elaborated user-goal and the elaborated new links. The link with the highest cosine value is the link predicted by CoLiDeS + Pic.

Simulation Study

A pilot study with a mock website on Human Body was conducted to verify the accuracy of CoLiDeS + Pic. The website had 4 levels of depth. We designed 8 user-goals, 2 for each level. We present here only the results of a goal concerning information on level3. Would there be any difference in predicting the right link after including features from a picture that is semantically close to the context? And, also what happens when we include a picture that is not so relevant?

Step1: Selecting Semantic Space; LSA provides many semantic spaces to facilitate representing accurately the background knowledge and general reading ability of different target user groups. We did choose the semantic space – “General Reading upto 1st Year College” (<http://lsa.colorado.edu/>)

Step2: User goal; User goal at level3 was – “*In the respiratory system, what name is given to the valve that drops down when we swallow in order to protect our lungs and trachea?*”

Step3: Elaborating Goal and Hyperlinks; LSA analysis is used to simulate the process of elaboration that happens during the comprehension phase of CoLiDeS. The goal mentioned in Step2 was elaborated as – “Larynx serves an

important biological purpose in addition to aiding in speech - to prevent foreign substances from entering the windpipe while swallowing, to forcefully expel foreign substances by coughing etc. Hyoid bone is horseshoe shaped and is the only bone in the body that floats, unconnected to another bone. A ring shaped cartilage is connected to the windpipe. The fat pad provides cushion to all the organs. What name is given to the organ present in our throat that drops down when we swallow in order to protect our lungs and trachea?” Similarly the hyperlinks were elaborated.

Step4: Elaborating features from pictures; Semantic features for five pictures were collected for this context from ten participants. Table 1 shows the five most frequent features obtained for the picture in Figure 4. Two features satisfying our criterion are shown in bold. Similarly, features representing other four pictures were selected. The features obtained for each picture were elaborated. High and low relevant pictures were then computed.

Table1: Most frequent features

| Feature | Frequency |
|----------------------------|-----------|
| Larynx, Vocal folds | 7 |
| Respiratory system | 6 |
| Wind pipe | 4 |
| Some bone | 3 |
| Front view and Back view | 3 |

Step5 and 6: Computation of Final Cosine Values; For each level, using LSA, cosine values were computed in the three conditions:

- 1) Elaborated Goal and Elaborated Links – Simple CoLiDeS (Step5)
- 2) Elaborated Goal and Elaborated Links appended with Elaborated Features of High Relevant Picture – CoLiDeS + High Pic (Step6)
- 3) Elaborated Goal and Elaborated Links appended with Elaborated Features of Low Relevant Picture – CoLiDeS + Low Pic (Step6)

Table 2 shows the final cosine values computed between the elaborated goal and the elaborated links appended with elaborated features in the three conditions. The correct links (Link1 for level1, Link1.1 for level2 and Link1.1.3 for level3) and the predicted links are shown in bold. We can see from the table that the correct link is predicted rightly by CoLiDeS and CoLiDeS + High Pic methods for all three levels. Furthermore, the cosine values obtained from CoLiDeS + High Pic are clearly higher than those obtained from simple CoLiDeS. CoLiDeS + Low Pic predicts the right link only at level 2. Moreover, the cosine values obtained from CoLiDeS + Low Pic are lower than those of

simple CoLiDeS. Similar computations for the other goals were conducted and they showed the same pattern.

Table 2: Final Cosine Values in the 3 Conditions

| Link | CoLiDeS | CoLiDeS + High Pic | CoLiDeS + Low Pic |
|------------------|-------------|--------------------------|-------------------------|
| Level1 | | | |
| Link1 | 0.19 | 0.29 | 0.09 |
| Link2 | 0.01 | 0.23 | 0.03 |
| Link3 | 0.16 | 0.26 | 0.13 |
| Link4 | 0.01 | 0.24 | 0.04 |
| Level 2 | | | |
| Link1.1 | 0.22 | 0.29 | 0.19 |
| Link1.2 | 0.18 | 0.27 | 0.08 |
| Level 3 | | | |
| Link1.1.1 | 0.19 | 0.26 | 0.12 |
| Link1.1.2 | 0.18 | 0.27 | 0.17 |
| Link1.1.3 | 0.32 | 0.46 | 0.12 |

Discussion & General Conclusion

Our first study showed that graphical information is equally important as textual information, and consequently should be incorporated in cognitive modelling of web-navigation. The second study in which we incorporated semantic features from pictures into CoLiDeS is giving higher cosine values in the highly relevant picture condition. Furthermore the results of the simulation show that if the information from pictures is included into modelling of navigation behaviour, the correct links are predicted with greater information scent. And most importantly, CoLiDeS + Pic would lead more frequently and more clearly to the right navigation path than the other conditions. We are also planning to corroborate these results with actual user behaviour in terms of number of clicks predicted by both the models. Whether the presence of a relevant picture decreases the perceived disorientation and the time taken to finish the task is to be tested. Finally, more computations with stricter conditions will be done to refine the procedure further.

References

- Blackmon, M.H., Polson, P.G., Kitajima, M., & Lewis, C. (2002). Cognitive Walkthrough for the Web. *2002 ACM conference on human factors in computing systems (CHI2002)*, 463-470.
- Blackmon, M.H., Kitajima, M., & Polson, P.G. (2005). Tool for Accurately Predicting Website Navigation Problems, Non-Problems, Problem Severity, and Effectiveness of Repairs. *2005 ACM conference on human factors in computing systems*. 31-40.
- Chi, E. H., Pirolli, P., Chen, K., & Pitkow, J. (2001). Using information scent to model user information needs and actions and the Web. *Proceedings of CHI 2001*, ACM Press, 490-497.
- Chi, E., Pirolli, P., & Pitkow, J. (2000). The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a website. *Proceedings of CHI 2000*, ACM Press, 161-168.
- Desimone, R., & Duncan, J., (1995), Neural Mechanisms of Selective Visual Attention, *Annual Review of Neuroscience*, March 1995, Vol. 18, 193-222.
- Hinesley, G.A. (2005). The impact of graphical conventions and layout location on search for webpage widgets. Unpublished Dissertation, University of Colorado, Boulder.
- Hinesley, G.A., & Blackmon, M.H. (2008). The Impact of Graphics and Location Expectations on the Search for Webpage Widgets. *Workshop on Cognition and the Web*, Granada, Spain.
- Itti, L., & Koch, C. Computational Modelling of Visual Attention, *Nature Reviews Neuroscience*, 2(3)194-203.
- Juvina, I., Oostendorp, H. van, Karbor, P., & Pauw, B. (2005). Toward Modeling Contextual Information in Web Navigation. *XXVII Annual Conference of the Cognitive Science Society*, Stresa, Italy.
- Juvina, I. & Oostendorp, H. van (2008). Modeling Semantic and Structural Knowledge in Web Navigation. *Discourse Processes*, 45(4-5), 346-364.
- Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. 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 CHI2000*, ACM Press, 357-373.
- Landauer, T. K., Foltz, P.W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.
- Mayer, R. E. & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38, 43-52.
- Miller, C. S., & Remington, R. W. (2004). Modelling Information Navigation: Implications for Information Architecture. *Human-Computer Interaction*, 19(3), 225-271.
- Oostendorp, H. van, & Juvina, I. (2007). Using a Cognitive Model to generate Web Navigation support, *International Journal of Human Computer Studies*, Volume 65(10), 887-897.
- Paivio, A (1986). Mental representations: a dual coding approach. Oxford. England: Oxford University Press.
- Pirolli, P., & Card, S.K. (1999). Information Foraging. *Psychological Review*, 106(4), 643-675.
- Pirolli, P., & Fu, W.T. (2003). SNIF-ACT: a model of information foraging on the World Wide Web. *9th International Conference on User Modeling (UM 2003)*; Johnstown; PA. Berlin: Springer Verlag; LNCS 2702: 45-54.