Gwizdka, J. (2008). Cognitive Load on Web Search Tasks. In Proceedings of Workshop on Cognition and the Web. Granada April 24-26, 2008. 83-86.

Cognitive Load on Web Search Tasks

Jacek Gwizdka, Dept. of Library & Information Science, Rutgers University, New Brunswick, NJ, USA
webcog2008 AT gwizdka.com

Abstract: Assessing cognitive load on web search is useful in examining which search system features, search tasks types and phases are more cognitively difficult for the searchers. It is also helpful in examining how individual differences among searchers (e.g. cognitive abilities) affect the search process. This investigation examined cognitive load from the perspective of primary and secondary task performance. A controlled web search study was conducted with 48 participants. The primary task performance components were found to be significantly related to the subjective task difficulty. However, the relationship between subjective task difficulty and the secondary task performance measures was weaker than expected. The results indicate that the dual-task approach needs to be used with caution.

Introduction and Background

Understanding factors that contribute to user's cognitive load on search tasks is crucial to identifying search system features, search tasks types and phases that impose too much load on users. For example, as new interactive features are introduced into the information search systems we need to understand why they are not widely used by users in spite of a strong support indicating their usefulness. User relevance feedback is one such example, where users have been reported to avoid providing relevance feedback due to the heightened cognitive load (Back & Oppenheim, 2001). Among other factors affecting search performance are the user's cognitive characteristics (e.g., Kim, 2004; Juvina & Oostendorp, 2004).

Methods used to date in assessing cognitive load included searcher observation, self-reports (e.g., using questionnaires, think-aloud protocols, and post-search interviews), dual-task techniques (Bruza et al., 1998; Kaki, 2005; Kim & Rieh, 2005), and various approaches that employ external devices to collect additional data on users (e.g., eye-tracking, pressure-sensitive mouse and other physiological sensors (Ikehara & Crosby, 2005)). The two latter groups of techniques have the advantage of enabling real-time, on-task data collection. However, use of external devices can be expensive and impractical. Hence, the promise of dual-task (DT) method that allows for an indirect objective assessment of effort on the primary task. A few studies employed this method to assess cognitive load in online search task. This research examines further the application of dual-task technique to the assessment of cognitive load on web search tasks.

Research Objectives

The current study extends our previous work (Gwizdka & Spence, 2006; Gwizdka & Spence, 2007) that examined factors affecting subjective perception of search task difficulty. We aim to understand better cognitive load on performance of search tasks. The current study examined dual-task method as an assessment technique of cognitive load on search tasks and considered the effects of selected individual differences and the contribution of different types of "cognitive actions" (e.g., query formulation, search results inspection, reading individual web pages, relevance judgment) to the searcher's perception of task difficulty. In particular, the study was conducted to examine:

- relationships between the searcher's "cognitive activities" and subjective perception of task difficulty;
- real-time assessment of cognitive load by employing dual task methodology;
- whether the selected cognitive abilities affect search and dual task performance.

Method

Forty eight participants (17 females; mean age 27 years) participated in question-driven, web-based information search study conducted in a controlled experimental setting. Participants were recruited from Rutgers University student population (undergraduate and graduate).

User Tasks

The study search tasks were motivated by questions that described what information needed to be found and provided a context for the search. The tasks were designed to differ in terms of their difficulty and structure. Twelve questions were used in total, eight out of which were created by Toms et al. (2007), while four simple fact-finding tasks were created for this study. Two types of search tasks were used: Fact Finding (FF) and Information Gathering (IG). The goal of a fact finding task is to find one or more specific pieces of information (e.g., name of a person or an organization, product information, a numerical value; a date). The goal of a information gathering task is to collect several pieces of information about a given topic. The tasks were also divided into three categories that depended on the structure of the underlying information need, 1) Simple (S), where the information need is satisfied by a single piece of information (by definition, simple task is of fact finding type); 2) Hierarchical (H), where the information need is satisfied by finding multiple characteristics of a single concept (a depth search); 3) Parallel (P), where the information need is satisfied by finding multiple concepts that exist at the same level in a conceptual hierarchy (a breadth search) (Toms et al., 2007). By definition, there were five possible combinations of task types and structure: FF-S, FF-H, FF-P, IG-H, and IG-P. We categorized tasks into three levels of "objective" difficulty based on their characteristics. FF-S was assigned low difficulty level, FF-P and FF-H middle-difficulty level, and IG-H and IG-P high difficulty level. During the course of each study session, participant performed six tasks of differing type and structure (Table 1). For each task, participant was able to choose between two questions of the same type and structure but on different topics. We offered the choice to increase the likelihood of participants' interest in the questions topic. The order of tasks was partially balanced with respect to the task type and structure (Table 1); it was not fully balanced due to the economics of the experimental study.

Table 1. Task rotations (for one rotation of search system).

Search System & UI	1			2		
QR / Task Seq.	TSeq1	TSeq2	TSeq3	TSeq4	TSeq5	TSeq6
QR1	FF-S1	FF-P1	IG-H1	FF-S2	FF-H1	IG-P1
QR2	IG-H1	FF-P1	FF-S1	IG-P1	FF-H1	FF-S2
QR3	FF-S1	FF-P1	IG-H1	IG-P1	FF-H1	FF-S2
QR4	IG-H1	FF-P1	FF-S1	FF-S2	FF-H1	IG-P1

A secondary task was introduced to obtain indirect objective measures of user's cognitive load on the primary search task. A small pop-up window was displayed at a fixed location on a computer screen at random time intervals. The pop-up contained a word with a color name (Figure 1). The color of the word's font either matched or did not match the name of the color. Participants' were asked to click on the pop-up as soon as they noticed it. The click was performed either by the right (match between the color name and the font color) or by the left mouse button (no-match). The use of color names and font colors (Stroop, 1935) in the DT was added to ensure cognitive engagement of users in the secondary task and to avoid automaticity.

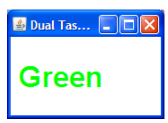

Figure 1. The secondary task pop-up window (not to scale).

Search System

The search tasks were performed on the English Wikipedia. Two different search systems and interfaces were employed (Google Wikipedia search, and ALVIS Wikipedia search (Buntine et. al. 2005)). The search interface was switched after task 3. The four task rotations were repeated for two orders of user interfaces. The analysis presented in this paper does not focus on the search system nor search interface features, and we do not discuss them in more detail.

Procedure

Each study session was an hour and a half to two hours long and was conducted in a university lab on a personal desktop computer running Microsoft Windows XP operating system. Each session consisted of the following steps: introduction to the study, consent form, three cognitive tasks (cognitive style w-a, mental rotation and operation span), search task practice, secondary task practice, background questionnaire, six search tasks, and post-session questionnaire. Before and after each search task, participants answered a short set of questions about their familiarity with and interest in the subject area, about subjective perception of task difficulty (before and after), about their search satisfaction. Web pages that searchers considered relevant were bookmarked and tagged by them. User interaction with computer (the primary and the secondary task events, visited and bookmarked URLs, mouse and keyboard events, and screen cam) was recorded using Morae software and the secondary task program.

Measures

Individual Difference measures (ID). We assessed two cognitive abilities, operation span (working memory performance) (Francis & Neath, 2003), and mental rotation (ability to manipulate mentally spatial images) (McGraw et al., 1999). We split the measures at median into high and low groups.

Cognitive action counts (CA). We recorded searcher's visits to web pages of various types. In particular, of our interest were counts of visits to pages related to "cognitive actions", these included entering queries and tags, making decisions what pages to examine and making relevance judgments.

Secondary task performance (DT). We recorded searcher's interactions with the secondary task (DT). The following measures were derived: 1) Number of missed DT events; 2) Total length of display time of the DT pop-up windows that were missed, its mean value per missed DT, and normalized to task duration; 3) Mean reaction time to DT events; 4) Ratio of correctly clicked to all clicked DT pop-ups; 5) Ratio of the estimated to the number of actual DT events. These measures were expected to reflect cognitive load on the primary task.

Subjective difficulty measure (SD). Upon the completion of all search tasks, searchers assessed difficulty of all tasks by ranking them on a 3 point scale (low-medium-high difficulty).

Results

We assume that observed cognitive actions (CA) on the primary task reflect cognitive effort of a searcher. Performance on the secondary task (DT) should reflect cognitive load on the primary task. Subjective perception of task difficulty reflects various components, including cognitive effort on a task. Therefore, SD, CA and DT measures should differ between the three levels of subjective task difficulty. In addition, cognitive abilities may affect both CA and DT. Our analysis of results was designed to examine these relationships (Figure 2).

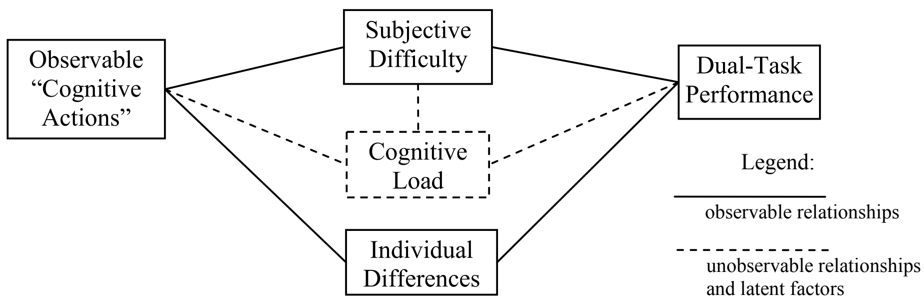

Figure 2. Model of examined relationships.

We examined how the CA and DT measures differed across the three levels of subjective difficulty. All individual CA variables differed significantly (Kruskal-Wallis test, χ^2=72.5, df=2, p<.0001). The total number of cognitive actions was for low subjective difficulty tasks 13.8 lower than for high difficulty tasks (post-hoc Bonferroni test, p<.001) and 8 lower than for medium difficulty (p<.001), while for medium difficulty tasks it was 5.8 lower than for high difficulty (p<.01).

We found that three DT measures differed significantly across the three levels of subjective task difficulty. The ratio of correct DT clicks to all DT clicks was 94% for low and medium difficulty tasks, while for 87% for high difficulty tasks. For low difficulty tasks, the number of DT events tended to be overestimated by 30%, for high and medium difficulty tasks the average number of estimated DT events was about right. The length of display time for missed DT pop-up window events was 15 seconds for low difficulty tasks, 21 seconds for medium, and 29 seconds for high; the difference between low and high difficulty tasks was significant (p<.05). However, there was no similar significant effect for the time of missed DT events normalized to the task duration. The differences in the values of the significant DT measures between the levels of task difficulty were in the expected direction. The above CA and DT relationships matched our expectations. However, the relationship between DT variables and subjective difficulty was weaker than expected.

To examine further whether measures of DT task performance are useful in assessing cognitive load on search tasks, we performed three regression analyses. Subjective task difficulty was the dependent variable, while independents (predictors) were as follows: 1) CA measures; 2) DT measures; 3) CA and DT measures combined. R^2 for the obtained models was .3**,** .05, and .31 respectively. Variables included in the model obtained with the combined set of predictors are shown Table 2. Clearly, CA measures are much stronger predictors of subjective task difficulty than DT measures.

We also examined whether the CA and DT variables were related to differences in cognitive abilities. Cognitive action measures were generally not related to the differences in cognitive abilities. In contrast, many DT measures differed significantly between the levels of working memory and spatial ability. For the two levels of the mental rotation ability, DT measures differed significantly (DT correctness and reaction time with p<.05, DT missed presence p<.001) in the expected direction; higher spatial ability was associated with better DT performance. However, for the two levels of working memory ability, the direction of the relationship was unexpected. Higher working memory ability was associated with lower levels of DT performance.

Table 2. Combined regression analysis. Predictors of subjective task difficulty (R^2=.31).

Variable Group	Variable	Stand. Beta coeff.	Incremental contrib. to variance explained
CA	Num. of individual results examined ****	-0.49	+23%
	Num. of first search result pages examined ****	-0.23	+3%
DT	Ratio of correct to all clicks on DT pop-up ***	0.13	+2%
CA	Num. of bookmarks **	0.17	+2%

**p < .05 ***p < .01 ****p < .001

Conclusions and Future Work

Work presented in this paper focused on examining dual-task method as an assessment technique of cognitive load and considered relationships between selected individual differences, "cognitive actions" and subjective search task difficulty. The modality of the primary task and the secondary task were the same (visual-manual), thus one could expect interference between the tasks and, as a result of central capacity limitations, a drop in performance on the secondary task with increased load on the search task, as well as a significant relationship between DT measures and subjective task difficulty. Lack of such significant relationships is likely to be due to the relative low difficulty of search task (approximately 50% of user/task instances were rated as low difficulty). Usefulness of the secondary task performance in assessment of cognitive load depends on task load. In conditions when the total load of primary and secondary tasks is not sufficiently high, the DT measures may not work well. The results indicate that the DT technique need to be used with caution.

We plan to explore more "sensitive" measures in subsequent experiments, to employ more difficult search tasks and possibly different user populations. The analysis presented in this paper was performed at the task granularity. In the future, we plan to examine how the cognitive load changes between different stages of a search task (e.g, query formulation, result list examination, individual document reading).

References

Back, J., & Oppenheim, C. (2001). A model of cognitive load for IR: Implications for user relevance feedback interaction. Information Research, 6(2).

Bruza, P., McArthur, R., and Dennis, S. (1998). Searching the world wide web made easy? The cognitive load imposed by query refinement mechanisms. In Proceedings of ADCS'98. pp. 65–71.

Buntine, W., Valtonen, K., Taylor, M. (2005). The ALVIS Document Model for a Semantic Search Engine. 2nd Annual European Semantic Web Conference in Heraklion, Crete May 29, 2005.

Francis, G., Neath, I. (2003). CogLab on a CD. 3rd Ed. Wadsworth Publishing.

Gwizdka, J., Spence, I. (2005). Indirect Assessment of Web Navigation Success. Extended Abstracts of ACM Conference on Human Factors in Computing Systems CHI'2005. ACM Press. 1427 - 1430.

Gwizdka, J., & Spence, I. (2006). What Can Searching Behavior Tell Us About the Difficulty of Information Tasks? A Study of Web Navigation. Paper presented at the ASIS&T 2006.

Gwizdka, J., & Spence, I. (2007). Implicit Measures of Lostness and Success in Web Navigation. Interacting with Computers, 19(3), 357-369.

Ikehara, C. S., & Crosby, M. E. (2005). Assessing cognitive load with physiological sensors. Paper presented at the Proceedings of HICSS'2005.

Juvina, I. & van Oostendorp, H. (2004). Individual Differences and Behavioral Aspects Involved in Modeling Web Navigation. Lecture Notes In Computer Science, 3196, 77-95.

Kaki, M. (2005). Findex: search result categories help users when document ranking fails. In Proceedings of CHI 2005. pp. 131–140, New York, NY, USA. ACM Press.

Kim, Y. M. & Rieh, S. Y. (2005). Dual-Task Performance as a Measure for Mental Effort in Library Searching and Web Searching. In Proceedings of the 68th Annual Meeting of ASIS&T.

Kim, K. S. (2004). Implications of User Characteristics in Information Seeking on the World Wide Web. International Journal of Human-Computer Interaction, 13(3), 323-340.

McGraw, K. O., Williams, J. E., Tew, M. D. (1999). PsychExps: A Web-based Cognitive Psychology Laboratory. WebNet 1999: 1348-1349.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18(6), 643-662.

Toms, E., MacKenzie, T., Jordan, C, O'Brien, H., Freund, L., Toze, S. Dawe, E., MacNutt A., (2007). How Task affects Information Search. Workshop Pre-Proceedings In Initiative for the Evaluation of XML Retrieval (INEX) 2007. N. Fuhr, N. Lalmas, A. Trotman Eds. http://inex.is.informatik.uni-duisburg.de/2007/inex07/pdf/2007-preproceedings.pdf [Accessed: January 19, 2008].