Gwizdka, J. (in press, 2008). Revisiting Search Task Difficulty: Behavioral and Individual Difference Measures. To appear in Proceedings of the 71th Annual Meeting of the American Society for Information Science and Technology (ASIST).

Revisiting Search Task Difficulty: Behavioral and Individual Difference Measures

Jacek Gwizdka

Department of Library and Information Science, Rutgers University
4 Huntington St, New Brunswick, NJ 08901, USA. assist08@gwizdka.com

Search task characteristics are recognized as important factors that affect search process and its outcomes. We examine the relationships among operational measures of searcher's behavior, individual cognitive differences, subjective task difficulty and mental effort assessed by dual-task performance. A web-based information study was conduced in a controlled experimental setting. Forty eight study participants performed six search tasks of varying type and structure. Subjective task difficulty was found to be influenced by the searcher's effort measured as the number of result pages and individual documents visited, the number of documents marked as relevant, as well as by individual cognitive differences, and mental effort assessed by performance on the secondary task. In contrast to previous studies, no strong effects of user navigation graph structure were found.

Introduction

Search task characteristics are recognized as important factors that affect search process and its outcomes. The importance of task characteristics has been recognized in Information Science early on. Task characteristics were seen as influential factors in information seeking behavior (Belkin, 1980) and their understanding was postulated to be crucial for improving the effectiveness of IR systems (Ingwersen, 1992).

In recent years, task became a focal point of many information studies. An increasing number of studies are concentrating on a specific characteristic of the task as an independent variable to explain associated information-seeking activities. A significant group among these studies has addressed task complexity or difficulty (for example, Byström, 2002; Byström & Järvelin, 1995; Gwizdka & Spence, 2006; Kim, 2006; Kim & Rieh, 2005; Toms et al., 2007; Vakkari, 1999).

Several researchers approached tasks and their effects in a comprehensive and systematic manner (Byström & Hansen, 2002; Kim & Soergel, 2005). A faceted classification of search tasks was proposed by Li (2004). Kim (2006) looked at factual, interpretive, and exploratory tasks and found that the level of task complexity was correlated with the search interaction. In our previous study (Gwizdka & Spence, 2006), we used nine tasks of varied complexity and found that the searchers' assessment of task difficulty was related to the number of unique web pages visited, time spent on each page, linearity of the searcher's navigation path, and its deviation from the optimal path. Most recently Li (2007) explored the relationships among work tasks, search tasks, and interactive search behavior. Work tasks facets were found to affect users' interaction with information systems. In particular, objective work task complexity affected almost all studied aspects of search behavior. Toms and her colleagues (Toms et al., 2007) investigated performance on search tasks that were complex by design. They examined effects of task type and structure on performance and users' perception of the task.

Research presented in this paper continues our prior work and examines the relationships among operational measures of searcher's behavior, individual cognitive differences, subjective task difficulty and mental effort assessed by dual-task performance.

Research Objectives

Study presented in this paper extends our previous research (Gwizdka & Spence, 2006; Gwizdka & Spence, 2007) by including new variables into the examination of factors that affect subjective assessment of search task difficulty. This study considered the effects of selected individual differences and the contribution of "cognitive actions" (expressed by the visits to different types of web pages, such as search result pages, individual results) to the perception of task difficulty. In particular, the study was conducted with the following goals in mind:

- to examine relationships between the searcher's activities and the subjective posterior perception of task difficulty;

- to examine which of the searcher's actions are better predictors of subjective task difficulty;
- to examine whether the relationship between the searcher's actions and post-task difficulty is affected by the levels of task variables;
- to examine whether the cognitive factors affect the subjective perception of search task difficulty;
- to examine relationship between mental effort and perception of search task difficulty.

Method

We conducted a question-driven, web-based information search study in a controlled experimental setting.

User Tasks

User search tasks were motivated by questions that described what information needed to be found and provided a context for the information need. The tasks were designed to differ in terms of their difficulty and structure. Twelve tasks were used in total, out of which eight were created by Toms et al. (2007) for INEX 2007 (Larsen et al. 2007), while four simple fact-finding tasks were created by us. 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 an 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 (an Information Gathering task cannot be Simple by definition).

We assigned a three-level "objective" difficulty rating to tasks based on their design. The simple fact finding tasks (FF-S) were assigned the easy level, the parallel and hierarchical fact finding (FF-P and FF-H) task were assigned the middle-difficulty level, and the information gathering tasks (IG-H and IG-P) were assigned the difficult level.

During 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 participant's interest in the task topic. The order of questions was partially balanced with respect to the task type and structure (Table 1). The order of questions was not fully balanced due to the economics of the study.

Table 1. Task rotations.

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 measures of user's mental effort (or cognitive load) on the primary search task (Kim & Rieh, 2006). A rectangle (a pop-up window presented in Figure 1) was displayed at a fixed location on a computer screen (top-right corner) at random time intervals (15-29 seconds) and for a random period of time (5-9 seconds). The length of a cycle was thus between 20 and 38 seconds. The rectangle contained a word with a color name (six well known colors were used). Participants were asked to position the mouse pointer within the pop-up rectangle and to click on it as soon as they noticed it. The color of font, in which the word was displayed, either matched or did not match the word. Depending whether they matched or not, participant was to click either the right (match) or the left (no-match) mouse button. The pop-up window disappeared as soon as it was clicked on. The use of color names and font colors in the secondary task was based on the Stroop effect (Stroop, 1935). The Stroop-like component was added to ensure cognitive engagement of users in the secondary task and to avoid automaticity (perceptual and motor reaction to a visual stimulus). The secondary task was controlled by a Java program created by us for this study.

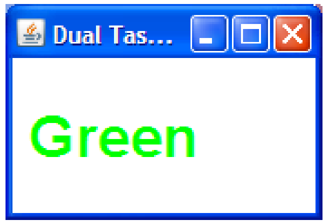

Figure 1. The secondary task pop-up window.

Search System

The search tasks were performed on the English Wikipedia. Two different search interfaces were employed. U1: Google Wikipedia search, and U2: ALVIS Wikipedia search (Buntine & Taylor, 2004; Luu et al., 2006). The search interface was switched after task 3. The four task rotations were repeated for two orders of user interfaces, U1/U2 and U2/U1. 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. However, the search system & interface (UI) was used as a factor in the analysis, and we report the significant findings.

Participants

Forty eight participants (17 females) participated in this study. The average participant age was 27 years. On the average, study participants reported that they searched internet several times a day and agreed that they were typically satisfied with their search results. Participants were recruited from Rutgers University student population (undergraduate and graduate) and received US$20 or a 5% course credit for their participation in the study. In addition, participants, who were most effective in their search task performance, received an additional bonus after the study was completed.

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 (demographics and internet use), six search tasks, and post-session questionnaire. Before and after each search task, participants answered a short set of questions (four questions before and six after a task) 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 marked and tagged by them. User interaction with computer (primary and secondary task events, visited and bookmarked URLs, user responses to the secondary task, mouse and keyboard events, and screen cam) was logged by using Morae software package and the secondary task Java program.

Measures & Variables

Controlled Task Factors

We controlled the search task type and structure levels as described above (Table 1).

Individual Difference Factors

Participants were tested for two cognitive abilities (operation span and mental rotation) and a cognitive style (wholist-analytic). The tests were administered on a computer (Table 2). The particular cognitive factors were selected as likely to affect searchers' performance on tasks (Chen et al., 2000; Chen et al., 2004; Gwizdka & Chignell, 2004).

Table 2. Cognitive tasks used in the study.

Cognitive Factor	Measures	Test Name	Short Description	Reference
Cognitive Style (wholist-analytic)	CSA_WA ratio. Wholistic style → 0 Analytic style → 2	Extended Cognitive Style Wholistic-Analytic Test (E-CSA) (from Department of Psychology, University of Edinburgh).	Cognitive style W-A reflects how individuals organize and structure information. Individuals described as Analytics tend to deconstruct information into its component parts, whereas individuals described as Wholists tend to retain a global or overall view of information.	(Peterson et al., 2005)
Operation Span	OpSpan ratio. 0-100%. Higher score→ higher ability	CogLab on CD (Wadsworth)	Operation Span is one of the measures of working memory performance. Operation span predicts verbal abilities and reading comprehension.	(Francis & Neath, 2003)
Mental Rotation	Mental Rotation Mean Reaction Time (RT) and Mental Rotation ratio of correct responses.	PsychExperiments (Dept. of Psychology, Mississippi University)	Mental Rotation is the ability to mentally manipulate spatial images.	(McGraw et al., 1999; McGraw & Tsai, 1993; Shepard & Metzler, 1971)

Other individual factors included participant's age, gender, first language, and their Web search experience.

Observable User Behavior: Search Effort, Search Efficiency and Action Speed

We used the logged time-stamped sequence of URLs to calculate measures of the searcher's behavior. The measures were calculated for each search task and included the number of web pages visited (TotalN); number of unique web pages visited (UniqueN); number of visits to pages with search results (a distinction was made between the first page of results (SrchResFirst) and the subsequent result pages (SrchResNxt)); number of individual results visited (IndRes); number of bookmarked individual result pages (BkmrkN); average time per web page (time_per_click); and the total time on each task (duration). These measures can be considered as belonging to two groups, *Search Effort* and *Action Speed* (Gwizdka & Spence, 2006). All page visit counts, except the TotalN, were calculated without revisits. Revisits were accounted for by calculating a separate measure described below.

We calculated three derived measures related to *Search Efficiency*.

1. Ratio of page revisits (Tauscher & Greenberg, 1997) was calculated using the ratio of unique (UniqueN) to all pages visited (TotalN) in the following way: revisit_ratio = 1 – UniqueN / TotalN. The higher the revisit ratio, the more pages were revisited. Hence, the less efficient the searcher was.
2. Ratio of the number of bookmarked pages (relevant results) to all pages visited. The higher the ratio, the more relevant pages per page visited were found. Hence, the more efficient the searcher was.
3. If we consider the individual web pages visited by searcher to be the nodes of a graph and the links actually followed by the searcher to be the graph edges, we can compute the graph properties, such as stratum (Botafogo et al., 1992). Stratum was used to characterize

searcher's behavior on web navigation tasks by McEneaney (2001), Shih et al. (2004), Herder et al. (2004), Juvina et al. (2004) and Gwizdka & Spence (2006). Stratum varies between zero and one. A value close to zero indicates a less linear navigation path; a value close to one indicates a more nearly linear navigation path.

Mental Effort

Mental effort was assessed by measuring the searcher's performance on the secondary task (DT). Five measures were derived:

1. Number of missed secondary task events;
2. The total length of display time of the secondary task pop-up windows that were missed (i.e. searchers did not click on them);
3. Mean reaction time to secondary task events;
4. Ratio of correctly clicked to all clicked secondary task windows;
5. Ratio of estimated to actual secondary task events (the estimation is performed after the main task is completed, thus it is a less direct measure).

Subjective Perception

Upon the completion of all search tasks, searchers assessed difficulty of all tasks by ranking them on a 3 point scale.

Results and Discussion

To gain a better understanding of the structure underlying the space of variables used in this study, principal component analysis (PCA) with varimax rotation was applied to the data. Due to space limitation, full results from this analysis cannot be presented in this paper. We note that the factors extracted in PCA corresponded roughly to the variable groupings described above (Search Effort, Search Efficiency, Action Speed, Mental Effort, Individual Difference Factors, Subjective Perception). This result supports our use of these groupings.

Searcher's Behavior and Subjective Post-Task Difficulty

Comparison with the Previous Predictive Model

The first objective was to find factors that affect searcher's perception of task difficulty. We started by examining regression model described in our ASIST 2006 paper (Gwizdka & Spence, 2006). The model stated that tasks tended to be a posteriori perceived as more difficult if more unique web pages were visited (UniqueN), more time was spent on each web page (time_per_click), and the navigation path was farther away from the optimal path (LCSlenMax), as well as it was less linear (-Stratum). Due to a different nature of information tasks in the current study, the notion of the optimal navigation path did not apply. Therefore, we included three variables: UniqueN, time_per_click, and Stratum. Only one variable continued to be a significant predictor: the number of unique web pages visited on a search task. The following model was obtained (shown with estimated b-coefficients):

$$\text{Predicted SubjDiff} = -.1*\text{UniqueN} + 4.9; \quad R^2=.28 \qquad (1)$$

Table 3. Predictors of subjective task difficulty (SubjDiff) in the first model.

Component group	Independent variable	Standardized Beta coeff.	Incremental contribution to variance explained
Search effort	Number of unique web pages visited (UniqueN)****	-.53	+28%

**** p < .001

The model can be interpreted by stating that tasks tended to be perceived a posteriori as more difficult if more unique web pages were visited (Figure 2). That may seem a simple result. A more interesting result is why other variables are absent from the current model. For example, linearity of the navigation path (measured by Stratum) was not significant in the current model. Our ASIST 2006 study used browsing tasks, while the current study used query-based search tasks that relied less on navigation. Other research that demonstrated significant relationship between task outcomes and the navigation path linearity also employed browsing tasks (McEneaney, 2001; Shih, et al., 2004). We conclude that Stratum is a better predictor of task difficulty for tasks that rely more on navigation. This conclusion supports and extends discussed by Gwizdka & Spence (2007) dependence of effects of Stratum on the nature of information search tasks.

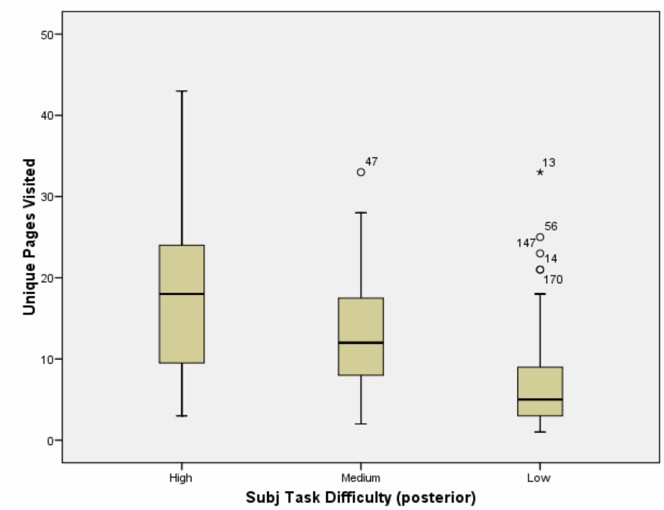

Figure 2. The number of unique pages visited by searchers shown for different levels of subjective task difficulty.

We also examined the same regression model for the data split by the task type, task structure and individual cognitive abilities. In all cases, no difference was found and the unique number of visited pages remained the significant predictor. However, R^2 was generally greater for searchers with higher cognitive abilities, than for those with lower abilities. For example for people low on Mental Rotation Correct R^2 was .19, while for those high R^2 was .37. Similarly, for low Operation Span (OpSpan) R^2 was .17, while for high R^2 was .35. Thus, search effort expressed as the number of visited pages may be a stronger predictor of subjectively perceived post-task difficulty for people, who are characterized by higher levels of cognitive abilities. T-tests were used to examine significance of the differences between high and low groups. The only significant difference in the number of unique pages was found for the different levels of Operation Span[1]. It was significant (at $p<.05$) with 9 web pages visited by searcher's low on OpSpan vs. 11 pages for high on OpSpan (Table 4). At the same time, participants who scored lower on OpSpan task tended to assess search tasks as difficult, while those who scored higher tended to assess search tasks as medium difficult ($F(1,282)=5.51$, $p<0.05$).

Table 4. High and low on Operation Span task.

Operation Span group	Subjective post-task difficulty SubjDiff	Number of unique pages visited UniqueN
low	high difficulty	9

[1] OpSpan was significantly correlated with the searcher's age, but the interaction effect of OpSpan*age on SubjDiff was insignificant.

high	medium difficulty	11

Apparently, for people with higher OpSpan, an increased search effort was not an important factor contributing to their perception of an increased task difficulty. Explanation of this surprising result is not straightforward and requires further research.

Extending the Predictive Model

Next, we examined relationship between subjective assessment of task difficulty and other variables that were measured in the current study. In particular, to answer what kinds of visited web pages contribute to the perception of task difficulty, we replaced in our regression analysis the unique number of all visited pages with the number of search result pages (first and subsequent), individual result pages, and bookmarked pages.

A non-parametric Kruskal-Wallis test for N-independent samples was used to examine if the variables differed for different levels of subjective task difficulty. Significant results are presented in Table 5, Table 6, and Table 7.

Table 5. Kruskal-Wallis tests for Search Effort measures and their mean values for SubjDiff.

	Task Duration (s)	**Unique Pages Visited**	**First Search Results Page**	**Beyond First Search Results Page**	**Individual Result (page)**	**Number of Bookmarks (relevant results)**
Chi-Square	67.8	81.1	51.6	17.7	67.7	15
df	2	2	2	2	2	2
Asymp. Sig.	0.0000	0.0000	0.0000	0.0001	0.0000	0.0006
SubjDiff	**Mean values of variables for three levels of subjective task difficulty (SubjDiff)**					
low	344	6.8	2.6	0	3.2	2.2
medium	557	13	4.2	0.1	6.4	3.1
high	764	18.1	5.9	1.3	8.1	3,3

The numbers of all individually examined types of web pages differed significantly between the levels of perceived task difficulty. For example (Table 5), the mean number of visited individual results was 3.2, 6.4 and 9.1 for low, medium and high difficulty tasks respectively (assessed significant by using Bonferroni test, p<.001). The mean number of results bookmarked b participants as relevant was for smaller by approximately 1 low difficulty tasks than for medium or high difficulty tasks (Bonferroni test, p<.05).

Table 6. Kruskal-Wallis tests for Search Efficiency measures.

	Ratio of revisits to pages	**Ratio of bookmarks to all visited pages**	**Stratum**
Chi-Square	10.462	40.447	26.951
df	2	2	2
Asymp. Sig.	0.0053	0.0000	0.0000

The mean ratio of pages judged as relevant to all pages visited by the searcher was the highest for low difficulty tasks (30%). For medium and high difficulty tasks, the mean ratio was 10% and 15%, respectively (Bonferroni test, p<.001). The Search Efficiency was clearly higher for tasks assessed by participants as easy.

Table 7. Kruskal-Wallis tests for Mental Effort and Individual Difference Factor measures.

	Presence of missed DTs [ms]	Ratio: user estimated DT / user clicks	Ratio correct to all clicks on DT	Cognitive Style WA
Chi-Square	6.5	12.9	8.2	7.4
df	2	2	2	2
Asymp. Sig.	0.04	0.002	0.02	0.025

The length of display time for missed DT window events was 15 seconds for low difficulty tasks, 21 seconds for medium, and 29 seconds for high. The display time difference between low and high difficulty tasks was significant (Bonferroni test, $p<.05$). 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 (these differences were significant according to the Kruskal-Wallis test, but insignificant according to post-hoc Bonferroni test). For low difficulty tasks the number of DT events tended to be overestimated by 30%, while for high and medium difficulty tasks the average number of estimated DT events was about right (these differences were significant according to Kruskal-Wallis test, but insignificant according to Bonferroni test). The differences in the values of the significant Mental Effort measures between the levels of task difficulty were in the expected direction.

We used regression to find a set of variables that best predicted subjective task difficulty, and to establish the relative importance of the variables. The following linear model was obtained (shown with estimated b-coefficients):

$$SubjDiff = -.11*IndRes - .07*SrchRes1st - .53*UI - .16*SrchResNxt + 1.6*rBkmrkToAll + .82*rCorrToClicks + .45*CogStyle + 3.7; \quad R^2=.35 \qquad (2)$$

Table 8. Predictors of subjective task difficulty (SubjDiff) in the first model.

Group	Variable	Stand. Beta coeff.	Incremental contrib. to variance explained	Unique variance explained (type II SS)	Variable order in the model
Search Effort	Num. of individual results examined (IndRes)****	-0.31	+22%	6.7%	1
	Num. of first search result pages examined (SrchRes1st)****	-0.13	+4%	1.0%	2
	Num. of following search result pages examined (SrchResNxt)**	-0.11	+1.1%	1.1%	7
Search System	Search system & interface (UI)***	-0.19	+2.1%	3.2%	3
Search Efficiency	Ratio of the relevant results found to all visited pages (rBkmrkToAll)***	0.18	+2.3%	2.4%	4
Mental Effort	Ratio of correct to all clicks on DT (rCorrToClicks)***	0.12	+1.7%	1.4%	5
Cognitive Factors	Cognitive style (CogStyleWA)**	0.12	+1.4%	1.4%	6

**$p < .05$ ***$p < .01$ ****$p < .001$,

According to the obtained model, tasks tended to be perceived a posteriori as more difficult when more search effort was expended (more individual result pages, first search results pages, more subsequent search result pages were visited), when the search was less efficient (lower ratio of relevant results to all pages visited), and when the second search system was used. The relationship to the search system can be explained by noting that the second system (ALVIS) was less familiar to users, while all of them were familiar with the first search system

(Google). The last predictor in Table 8 is Cognitive Style WA. The direction of the relationship indicates that searchers, who tend to have analytic style are more likely to perceive tasks as easy.

We examined how the set of variables that predicted subjective task difficulty changed for different levels of objective task difficulty, task structure, and searcher's age.

1. *Objective task difficulty* (task type combined with structure). We obtained three models for each of the three levels of this variable. Search Effort and individual factors were represented in all three models. Cognitive style W-A affected perception of difficulty of the simple fact-finding tasks (FF-S). The most complex model was obtained for the difficult tasks (information gathering tasks – IG). That model contained search efficiency and user interface components, but also included in the model was Operation Span (searchers with higher OpSpan tended to perceive tasks as easier).
2. *Task structure*. All models for the three levels of task structure contained Search Effort. In addition, the model for simple tasks (FF-S) contained cognitive style, while the models for H and P tasks contained Search Efficiency. This result confirms our intuition that Search Efficiency plays a less important role for simple tasks, where the goal is to find one piece of information.
3. *Age*. We split the participants into two age groups at the median age (23 years). The obtained two regression models contained fewer variables than the main model. The Search Effort and Individual difference factors were represented in both models. For younger participants the cognitive style W-A played a significant role, while it was absent from the older participants' model, which included mental rotation ability. Older searchers, who had higher mental rotation ability tended to perceive tasks as easier. The effect of Search Efficiency appeared only for older participants. Search Efficiency was represented in this model as the absolute number of marked relevant results.

Dual-task performance and Subjective Post-Task Difficulty

We measured user's performance on the secondary task to derive measures of Mental Effort involved in the performance on the primary search task. We found that three of them differed significantly between the three levels of subjective task difficulty. One of them was a significant predictor (Table 8). To further examine whether the secondary task performance is useful in assessing Mental Effort on search tasks we analyzed how it is influenced by the task type and structure. The length of display time of missed DT events was the significant Mental Effort variable. The missed DT events were on the average displayed for 13 second longer during IG tasks than during FF-S tasks (Bonferroni test with $p<.01$), and 11 seconds longer during hierarchical than simple tasks (Bonferroni test with $p<.05$). Although, we found this significant result, the Mental Effort measures were somewhat less sensitive than expected. We should note that the analyses presented in this paper are per search task. Kim & Rieh (2005) suggested that dual-task method of assessing mental effort is useful for comparing differences in effort required by various stages of the search process. We need to perform further analyses to examine such use of the DT method.

Summary and Conclusions

This research aims to contribute to better understanding which search task attributes and factors affect search task difficulty. Work presented in this paper focused on examining the relationships among measures of searcher's behavior, individual cognitive differences, subjective task difficulty and mental effort assed by the dual-task method.

Subjectively assessed post-task difficulty was found to be influenced by the searcher's effort measured as the number of result pages and individual documents visited, the number of documents bookmarked as relevant, as well as by performance on the secondary-task and individual cognitive differences. The significant effects of variables on subjective task difficulty were influenced by task variables (task type and structure) and individual differences between searchers (e.g., age). However, only variables representing Search Effort were present in all regression models. In particular, variables that expressed the number of visited web pages (of various types) seem to be one of the most robust factors that influence searchers' perception of task difficulty. Hence, the general predictive models will likely contain higher level factors rather then specific individual variables.

As expected, individual differences in cognitive characteristics of searchers were found to influence their perception of task difficulty. The effects were smaller than expected, which may be due to a relatively small size of participant population. A couple of effects appeared only within specific levels of task or user factors, for example, for most complex tasks (IG) or for older participants (> 23 years). As expected, the direction of these effects indicated that individuals, who had higher abilities tended to perceive search tasks as easier.

One limitation of our study is due to a relatively large proportion of tasks assessed by participants as easy. In over 50% of cases, tasks were assessed as easy and only 15% of tasks as difficult. The caveat is thus to use search tasks of appropriate complexity. Toms et al., (2007) postulated to study complex tasks. To make tasks more difficult, one could, of course, design tasks to be more complex or one could study a different participant population (e.g., older, or less experienced people).

Characterization of information search tasks that are performed in different contexts and by different groups of people holds a promise for the creation of personalized information retrieval systems. For example, modeling tasks in terms of operational measures of the searcher's activities should ultimately enable us to build real-time predictive task models and, thus, to create information systems that adapt to the user. To be successful, however, such systems would need to be able to understand user tasks, and to assess the user's characteristics from their activities. We hope that our work brings us a step closer to this future goal.

References

Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, 5, 133-143.

Botafogo, R. A., Rivlin, E., & Shneiderman, B. (1992). Structural analysis of hypertexts: Identifying hierarchies and useful metrics. *ACM Transactions on Information Systems*, 10, 142-180.

Buntine, W., Taylor, M. (2004). ALVIS: Superpeer Semantic Search Engine. European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies on November 25-26, 2004, Royal Statistical Society, London.

Byström, K. & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing and Management*, 31, 191-213.

Byström, K. (2002). Information and information sources in tasks of varying complexity. *Journal of the American Society for Information Science and Technology*, 53, 581-591.

Byström, K. & Hansen, P. (2002). Work tasks as unit for analysis in information seeking and retrieval studies. *Proceedings of the 4th International Conference on Conceptions of Library & Information Science* (CoLIS4), 239-252.

Chen, C. M., Czerwinski, M., & Macredie, R. (2000). Individual differences in virtual environments - Introduction and overview. Journal of the American Society for Information Science, 51(6), 499-507.

Chen, S. Y., Magoulas, G. D., & Macredie, R. D. (2004). Cognitive styles and users' responses to structured information representation. International Journal on Digital Libraries, V4(2), 93-107.

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

Gwizdka, J., Spence, I., (2005a). Predicting Outcomes of Web Navigation. *Poster presented at WWW'2005*, Chiba, Japan.

Gwizdka, J., Spence, I. (2005b). 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.

Gwizdka, J., & Chignell, M. (2004). Individual differences and task-based user interface evaluation: A case study of pending tasks in email. Interacting with Computers, 16(4), 769-797.

Hansen, P. (1999). User interface design for IR interaction: A task-oriented approach. *Proceedings of the Third International Conference on Conceptions of Library and Information Science (CoLIS 3).* 191-205.

Herder, E. (2003). Revisitation Patterns and Disorientation. *Proceedings of the German Workshop on Adaptivity and User Modeling in Interactive Systems ABIS 2003*, 291-294.

Ingwersen, P. (1992). *Information Retrieval Interaction*.

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

Kim, J.-H. (2006). Task as a Predictable Indicator for Information Seeking Behavior on the Web. *An unpublished dissertation*. SCILS, Rutgers University.

Kim, Y. M. & Rieh, S. Y. (2005). Dual-Task Performance as a Measure for Mental Effort in Library Searching and Web Searching. *Proceedings of the 68th Annual Meeting of the American Society for Information Science and Technology*.

Kim, S. & Soergel, D. (2005). Selecting and Measuring Task Characteristics as Independent Variables. *Proceedings of the 68th Annual Meeting of the American Society for Information Science and Technology*.

Larsen, B., Tombros, T. & Malik, S. (2007). Interactive Track Guidelines for the Initiative for the Evaluation of XML Retrieval (INEX) 2007.

Li, Y. (2004). Task type and a faceted classification of task. Proceedings of American Society for Information Science and Technology 2004, November 13-17, RI: Providence, USA. [Retrieved Jan 2, 2008, from http://www.asis.org/Conferences/AM04/posters/250.doc]

Li, Y. (2007). Work Tasks, Search Tasks, And Interactive Information Searching Behavior: An Exploration Of Their Relationships. Unpublished dissertation. SCILS, Rutgers University

Luu, T., Klemm, F., Podnar, I., Rajaman, M., Aberer, K. (2006). ALVIS Peers: A Scalable Full-text Peer-toPeer Retrieval Engine. Workshop on Information Retrieval in Peer-to-Peer Networks (P2PIR 2006), collocated with the AMC Fifteenth Conference on Information and Knowledge Management (CIKM 2006), Arlington, VA, USA, November 11, 2006

McEneaney, J. E. (2001). Graphic and numerical methods to assess navigation in hypertext. *International Journal of Human Computer Studies*, 55, 761-786.

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

McGraw, K. O., & Tsai, W. S. (1993). A Computer Program for Conducting Mental Rotation Experiments. Teaching of Psychology, 20(4), 246-247.

Peterson, E. R., Deary, I. J., & Austin, E. J. (2005). A new measure of Verbal-Imagery Cognitive Style: VICS. Personality and Individual Differences, 38(6), 1269-1281.

Shepard, R.N. & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. Science (171) 701-703.

Shih, P.-C., Mate, R., Sanchez, F., & Munoz, D. (2004). Quantifying user-navigation patterns: a methodology proposal. In Poster presented at the 28th International Congress of Psychology in Bejing Bejing 2004.

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

Tauscher, L. & Greenberg, S. (1997). How people revisit web pages: Empirical findings and implications for the design of history systems. *International Journal of Human Computer Studies*, 47, 97-137.

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].

Vakkari, P. (1999). Task complexity, problem structure and information actions. Integrating studies on information seeking and retrieval. *Information Processing and Management*, 35, 819-837.

Appendix.

Table 9. Selected search tasks used in the study (one for each combination of task type and structure).

Type	Question text
FF-S	You love history and, in particular, you are interested in the Teutonic Order (Teutonic Knights). You have read about their period of power, and now you want to learn more about their decline. What year was the Order defeated in a battle by a Polish-Lithuanian army?
FF-H	A friend has just sent an email from an Internet café in the southern USA where she is on a hiking trip. She tells you that she has just stepped into an anthill of small red ants and has a large number of painful bites on her leg. She wants to know what species of ants they are likely to be, how dangerous they are and what she can do about the bites. What will you tell her?
FF-P	As a history buff, you have heard of the quiet revolution, the peaceful revolution and the velvet revolution. For a skill-testing question to win an iPod you have been asked how they differ from the April 19th revolution.
IG-H	You recently heard about the book "Fast Food Nation," and it has really influenced the way you think about your diet. You note in particular the amount and types of food additives contained in the things that you eat every day. Now you want to understand which food additives pose a risk to your physical health, and are likely to be listed on grocery store labels.
IG-P	Friends are planning to build a new house and have heard that using solar energy panels for heating can save a lot of money. Since they do not know anything about home heating and the issues involved, they have asked for your help. You are uncertain as well, and do some research to identify some issues that need to be considered in deciding between more conventional methods of home heating and solar panels.

Acknowledgments

Thanks to study participants, and to Chang Liu, Jingjing Liu, Irene Lopatovska, and Will Tai, who conducted the experimental sessions. This research was supported in part by IMLS grant LG-06-07-0105-07.