Multi-Aspect Information Use Task Performance: The Roles of Topic Knowledge, Task Structure, and Task Stage

Jingjing Liu  
School of Library and Information Science  
University of South Carolina  
jingjing@sc.edu

Nicholas J. Belkin  
School of Communication & Information  
Rutgers University  
belkin@rutgers.edu

ABSTRACT
Information search, quite often, is not an isolated activity, but is accompanied by the use of the located information to generate some outcome. Frequently seen are “complex” tasks that consist of multiple aspects and can be divided into sub-tasks and/or finished in multiple sessions. This paper explores how search systems may help users with their multi-aspect tasks by examining whether, and how user knowledge and task structure play roles in leading to better work task performance. A 3-session lab experiment was conducted with 24 participants, each coming 3 times to work on 3 subtasks of a general task, couched either as the “parallel” or the “dependent” structure type. The overall task was to write a report on the general topic, with interim documents produced for each subtask. Results show that users’ previous knowledge of task topics positively correlated with task performance, but the evolving and the exit knowledge did not. Neither task session nor task structure significantly affected session task performance. However, there was a tendency that those users with different levels of task knowledge performed differently in different tasks: higher knowledge users tended to perform better in the parallel task than in the dependent; lower knowledge users tended to perform better in the dependent task than in the parallel. These results can possibly be explained by users’ searching and writing behaviors. Our findings help understand factors influencing information use task performance, and have implications on designing personalized systems that support information use task accomplishment.

Keywords  
Information use task performance, topic knowledge, multi-aspect task, task structure.

INTRODUCTION
Information retrieval (IR) systems target helping people find information. However, in everyday life, searching for information is often driven by motivating goals, such as to accomplish some work task at hand (Ingwersen & Järvelin, 2005; Li & Belkin, 2008). These tasks generally involve using the information besides finding it, for example, writing a research paper, or preparing a vacation plan. The fact that searching and using information are frequently interconnected calls for research on information behaviors in the general context of task accomplishment instead of search only. Previous studies have found that some measures that are typically used in evaluating search performance were not correlated with the performance of report writing (Liu & Belkin, 2012), and the effort in the search process degraded search precision but improved the essay writing task outcome (Vakkari & Huuskonen, 2012)

It is also seen in everyday life that work tasks often cannot be finished at once, but they may be divided into subtasks and spread over multiple sessions. Donato, Bonchi, Chi, & Maarek (2010) reported that 25% of the overall query volume on the web corresponds to multi-session tasks. It is likely that a good portion of the complex tasks is multi-aspectual. The relationships between the multiple aspects, or the subtasks, can be different. For example, Toms et al. (2007) designed their tasks in two structures, the parallel and the hierarchical. The former type consists of multiple concepts that exist on the same level in a conceptual hierarchy, and the latter consists of concepts that are on various, hierarchically-related levels. They found in their study that task structure affected users’ search behaviors. It is undiscovered though, if and how task structure may affect users’ work task performance.

As an important characteristic affecting information seeking behavior, user knowledge has received extensive research attention. Knowledge has been found to affect search behaviors and search performance. It remains unknown if, and how, knowledge affects work task performance, especially in multi-aspect information tasks.

The research reported here attempted to answer the following two research questions (RQs):

RQ1. Do users with different levels of topic knowledge perform multi-aspect tasks equally well?

RQ2. Does task structure play a role in multi-aspect task performance for users with different levels of knowledge?

LITERATURE REVIEW

Work Task and Its Performance
In people’s daily lives, searching for information is usually driven by some motivating goals or work tasks (Ingwersen

ASIS&T 2014, October 31-November 4, 2014, Seattle, WA, USA.  
Copyright to authors.
& Järvelin, 2005; Li & Belkin, 2008). The tasks can vary from planning for a trip, or shopping, to research, work reports, and so on. These tasks often require certain task outcomes, being reports, email responses, or essays/papers. While search performance has been consistently evaluated in the field of IR, much less effort has been spent on examining information use task performance. Evaluating search systems on the work task level has gained increasing attention recently, and was noted as “where the biggest challenges remain” in IR system evaluation (Wilson et al., 2010, p. 74). The following reviews a few related studies.

Wilson, Andre, & Schraefel (2008) designed an interface feature, named Backward Highlighting (BH), which allows information searchers to see and utilize associations in columns to the left of a selection in a directional column browser like iTunes. Participants were asked to find and write about facts in a report. One evaluation method was the number of facts in the reports. It was found that the proposed BH feature increased discovery of facts. Kammerer et al. (2009) designed a social tagging search browser and examined its effectiveness in helping people learn. Participants were asked to write, besides collecting webpages, a short coherent summary addressing aspects in the question domain. The quality of the summaries was measured by pre-defined topic-specific criteria about topic coverage levels based on a scale. They found that users of the proposed interface generated better quality summaries. Wilson & Wilson (2013) compared different techniques to measure one’s learning through self-generated summaries during the process of exploratory search. The techniques included counting facts, judging topic coverage, and measuring the depth of learning. They found that there is not a one-size-fits-all measurement, but various measures worked in different conditions according to users’ levels of base knowledge, and the length of their summaries.

Recently, some studies have examined the relationship between information search and information use tasks. Through examination of users’ Web search and their use of the located information in report writing, Liu & Belkin (2012) found some measures that are typically used in evaluating search performance, such as the number of issued queries, task completion time, time on searching, etc., were not correlated with the performance of report writing. Report writing performance was affected by users’ knowledge of the task topics and their previous task experience, and it was positively correlated with users’ effectiveness in finding useful webpages and their time allocated to writing. Vakkari & Huukonen (2012) examined how the effort participants spent on searching is associated with their search performance and task outcome in an essay writing task. They found that effort in the search process degraded search precision but improved the essay writing task outcome. This indicates, as the authors suggested, that evaluation of search systems should be including measures for search process and task outcome in addition to the search result precision.

Knowledge and IR
There has been an extensive body of literature on the relationship between searchers’ knowledge levels and their search behaviors and performance. Two different types of knowledge have been studied: 1) the knowledge of a subject/domain, called domain knowledge, and 2) users’ familiarity with a search task topic, called topic knowledge. These two types of knowledge have been shown to have different effects on users’ search behavior (Zhang, Liu, & Cole, 2013). While many studies examined the effect of domain knowledge on users’ search (e.g., Wildemuth, 2005), what is relevant to the current paper is topic knowledge (used interchangeably with topic familiarity), as reviewed below. We do not consider here knowledge of the search system itself, or of searching in general.

Hembrooke, Granka, Gay, and Liddy (2005) found that experts with high topic knowledge issued longer and more complex queries than novices. They also used elaboration as a reformulation strategy more often as compared to simple stemming and backtracking modifications used by novices. Allen (1991) found that compared with their counterparts, people with high topical knowledge used more search expressions, and they employed more search expressions that had not been contained in their statement of information need. Kelly and Cool (2002) found that with increasing familiarity with topics, reading time tended to decrease and the efficacy of search, measured by the ratio of the number of saved documents to the total number of viewed documents, increased. Kelly (2006) found that users’ topic familiarity, as a contextual factor, had a significant effect on user behaviors, specifically, document display time.

In the above studies, user knowledge was considered to be static. Some other studies considered differences in users’ search behaviors and performance as their knowledge levels changed. This complies with Kulthau’s (1991) Information Seeking Process (ISP) model, which points out that users’ cognitive states, i.e., thoughts, change during the search process. Using a quantitative study, Liu et al. (2013) confirmed the reasonable assumption that searchers’ knowledge increases during the information search and task completion process.

Information searching happens in a certain type of context and is for the purpose of accomplishing some work tasks or achieving some other motivating goals. To our knowledge, examining the relationship between work task performance and user knowledge has not been extensively studied. Acknowledging that people’s knowledge changes during their information search process, relationships should also be examined between evolving knowledge and task performance. This is also a gap in the literature.

Task Structure in IR
In examining the effects of tasks on information searchers’ behaviors and performance, a common approach is to classify user tasks into different types according to some task feature(s), for example: fact-finding vs. information
gathering (Toms et al., 2007; Kellar, Watters, & Shepherd, 2007). Li & Belkin (2008) constructed a comprehensive classification scheme that includes a number of dimensions of task features/attributes: task product, complexity, and difficulty, to name a few.

For complex tasks with multiple concepts or aspects, task structure becomes an important factor to look at, which we treat as one task feature in this article. Toms et al. (2007) designed different tasks, one feature of which was their conceptual structures. The two task types in their approach were: the parallel, where the search uses multiple concepts that exist on the same level in a conceptual hierarchy, and the hierarchical, where the search uses a single concept for which multiple attributes or characteristics are sought. They found that for most evaluation metrics, the hierarchical tasks required more user effort than the parallel ones: longer time processing the results of a query, more use of interface tools, and more saved items. Liu & Belkin (2010) looked at multi-session tasks, which were designed with parallel and dependent structures. The sub-tasks in the parallel task were parallel to each other, while in the dependent task, some subtasks were dependent upon the accomplishment of others. They found that task stage helped in the interpretation of document usefulness based on dwell time, but this role was different in the parallel and the dependent tasks.

**Task Stage in IR**

Search stage has been found to be a significant factor affecting search behaviors. Kulthau’s (1991) ISP (Information Seeking Process) model proposes that the information search process has six stages: initiation, selection, exploration, formulation, collection, and presentation. The user’s feelings, thoughts, and actions vary along the different stages. Vakkari and colleagues (e.g., Vakkari, 2001; Vakkari & Hakala, 2000) found that, as the task stage progressed, searchers’ vocabulary changed from broader to narrower terms, that they were less likely to start their initial queries by introducing all the search terms, were more likely to enter only a fraction of the terms, and tended to use more synonyms and parallel terms. White, Ruthven & Jose (2005) studied how the use and effectiveness of implicit relevance feedback (IRF) and explicit relevance feedback (ERF) is affected by three factors: search task complexity, the search experience of the user and the stage in the search. With respect to the search stage, their results suggest that IRF is used more in the middle of the search than at the beginning or end, whereas ERF is used more towards the end. Liu & Belkin (2010) explored whether dwell time can be a reliable indicator of document usefulness in multi-session tasks. They found that task stage helped interpret document usefulness from first dwell time, i.e., the first duration that a document was viewed.

**METHOD**

**Experimental Design**

A 3-session lab experiment was conducted, with 24 participants. The study was a between-subjects design, each participant assigned to one of two tasks, described below.

**Tasks**

Work tasks (simplified as “tasks”) in the study were designed to mimic journalists’ assignments since they could be relatively easily set as realistic tasks in different domains. Among the many dimensions of task types, this study focused on task structure, i.e., the inter-subtask relation. Tasks were designed with different sub-task structures while other facets in Li & Belkin’s (2008) task classification scheme were kept as constant as possible. This makes it reasonable to attribute observed differences to this single factor of task structure. Two task types were used in the study: the parallel and the dependent. This was similar to Toms et al. (2007), which classified tasks into parallel and hierarchical types according to the conceptual structures in the tasks. Both tasks in the current study had three subtasks, each of which was finished by the participant during one separate session, so the subtasks are also called session-tasks.

The general tasks asked the participants to write a three-section feature story on hybrid cars for a newspaper, and to finish and submit each article section at the end of each experiment session. They were told in the beginning that they needed to integrate the 3 sections into one article at the end of the 3rd session. In the dependent task, the three session-tasks were: 1) collect information on what manufacturers have hybrid cars; 2) select three models that you will mainly focus on in this feature story; and 3) compare the pros and cons of three models of hybrid cars. In the parallel task, the three session-tasks were finding information and writing a report on three models of cars from auto manufacturers renown for good warranties and fair maintenance costs: 1) Honda Civic hybrid; 2) Nissan Altima hybrid, and 3) Toyota Camry hybrid. It was hypothesized that the session-tasks in the parallel task were independent of one another, but in the dependent task, there would be perceived to be at least some notional order. To maintain consistency though, session-task orders in the task description in both tasks were rotated and users were allowed to choose whatever order of session-task performance they preferred. In the dependent task, most participants (8 out of 12) chose the above 1-2-3 order. Other orders were logical to the participants according to their comments in the exit questionnaire, keeping the task a dependent one to them.

In each session, participants were allowed to work up to 40 minutes to freely search on the web for helpful information and write and submit their reports. For logging purpose, users were asked to keep only one Internet Explorer (IE) window open and use back and forward buttons to move between web pages.
Participants
Since the tasks were journalists’ assignments, Journalism/Media Studies undergraduates were invited as participants in mimicking journalists. The 24 participants (21 female, 3 male) had an average age of 20.4 years. They were recruited via email to the student mailing list at the Journalism/Media Studies undergraduate program. Each participant came 3 times within a 2-week period based on his/her schedule, and most of them came the same day in three contiguous weeks. Each was assigned randomly to a task condition. Each obtained $30 payment upon finishing all 3 sessions. Participants were informed at the beginning of the experiment that the top 6 who submitted the most detailed reports would obtain an additional $20.

Procedure
Participants came individually to an information interaction lab for the experiment. Upon arrival in the first session, they completed a consent form and a background questionnaire. They were then given the general work task to be finished in the whole experiment. A pre-session task questionnaire collected their familiarity with the general task topic, previous experience with the type of task, and the expected task difficulty. Then they were asked to pick one session-task. A pre-session session-task questionnaire followed to collect their familiarity with the session-task topic, previous experience with the session-task, and expected session-task difficulty. Then they were given up to 40 minutes to search for information and write their report. Morae (http://www.techsmith.com/morae.asp) was used to record user-system interactions. After report submission, participants were asked to rate the usefulness of each document that they had viewed. A post-session session-task questionnaire and a post-session general task questionnaire were then administered to elicit user perceptions on the difficulty of collecting information for the task and session-task, degree of success with the submitted report, as well as satisfaction with the report.

In the 2nd and the 3rd sessions, participants went through the same steps except for the consent form and background questionnaire. In the 3rd session, after the post-session general task questionnaire, an exit interview asked them to reflect on their overall knowledge gain and to comment on the whole experiment.

In the experiments, all the ratings used 7-point Likert scales, where 1 was for the most negative; 7 was for the most positive, and 4 was the mid/neutral point.

Variables
Three groups of variables were examined in the study, which were about task performance, user knowledge, and users’ search performance/behaviors. Each included a number of variables, as introduced below.

Task Performance Measurements
As introduced before, participants were encouraged to write detailed reports by using the bonus $20 payment as an incentive. Therefore, the degree of detail is used as the criterion of task performance assessment. Similar to Wilson et al. (2008), that used the number of facts in the reports to assess users’ knowledge levels, the current study used the following to assess users’ report level of detail, counted by one of the authors:

- **Number of statements**: the number of sentences in each session’s report.
- **Number of facts**: the number of facts in each session’s report. For example, if a sentence talks about a car’s color, mileage per gallon, and engine size, these were counted as three facts.
- **Total number of statements**: total number of sentences in the final report (combination of the number of sentences in the three sessions).
- **Total number of facts**: total number of facts in the final report (combination of the number of sentences in the three sessions).

User Knowledge
Knowledge in this experiment was measured by self-rated degree of familiarity with task topics. This has been used in previous studies (e.g., Kelly & Cool, 2002; Liu et al., 2013). This study considers both the general and the session-task topic knowledge. The following variables were elicited in questionnaires:

- **Entry general task topic knowledge**: self-rated degree of familiarity with the general task topic measured in the pre-session questionnaire in the first session.
- **Evolving general task topic knowledge**: self-rated degree of familiarity with the general task topic measured in the pre-session questionnaire in each session. The purpose was to elicit participants’ evolving knowledge along the task completion process.
- **Exit general task topic knowledge**: self-rated degree of familiarity with the general task topic measured in the post-session questionnaire in the third (last) session.
- **Session-task topic knowledge**: self-rated degree of familiarity with the session-task topic elicited before each session.

User search performance/behaviors
To obtain an in-depth understanding of the relationships between task performance and other factors, a number of behavioral variables were examined in this research. These could also be viewed as ways to assess search performance. These variables were extracted from the logged data. The following lists the variables on a session basis:

- **Number of queries**: the number of queries in a session
- **Number of total pages**: number of all webpages viewed in a session
- **Number of unique pages**: number of unique webpages viewed in a session
- **Number of useful pages**: number of webpages viewed in a session with a usefulness rating score above 4 (somewhat useful)
• **Ratio of useful pages to all:** ratio of useful webpages out of all webpages viewed in a session
• **Number of unique useful pages:** number of unique webpages viewed in a session with a usefulness rating score above 4 (somewhat useful)
• **Ratio of unique useful pages to all:** ratio of unique useful webpages out of all webpages viewed in a session
• **Task completion time (in seconds):** total time that users spent on finishing the session-task. This combines both searching and writing times in a session.
• **Time on searching:** time spent on searching for information in a session
• **Time on writing:** time spent on writing report in a session
• **Ratio of searching time to all:** ratio of time spent on searching out of total task completion time in a session
• **Ratio of writing time to all:** ratio of time spent on writing out of total task completion time in a session

**RESULTS**

General Linear Model (GLM) (Haase, 2011) tests were conducted to gain a comprehensive understanding of the relationships among task type, task sessions, topic knowledge, and task performance. We chose GLM analysis because it can detect interaction effects between/among variables in addition to the main effects of independent variables on dependent variables. To detect the underlying reason, GLM tests were also conducted for the relationships among task type, task sessions, topic knowledge, and searching/writing time.

In order to have a comprehensive understanding of task performance, performance at both the general and the session-task levels was examined. Because the number of participants was rather small for reliable statistical analysis, we identify descriptive patterns as well as statistically significant results below.

**General Task Performance across Tasks**

As is shown in Table 1, there was no difference between the two tasks in either of the two overall task performance measurements: total number of statements and total number of facts. Those who worked with the parallel task and those with the dependent task finished them equally well, as assessed by our measurements for general task performance.

<table>
<thead>
<tr>
<th></th>
<th>Total number of statements Mean (SD)</th>
<th>Total number of facts Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Task</td>
<td>61.58 (20.14)</td>
<td>79.14 (35.64)</td>
</tr>
<tr>
<td>Parallel Task</td>
<td>56.83 (18.57)</td>
<td>84.33 (22.63)</td>
</tr>
<tr>
<td>Mann-Whitney U test U (p)</td>
<td>63.00 (.630)</td>
<td>91.00 (.291)</td>
</tr>
</tbody>
</table>

**General Task Performance and Users’ General Task Topic Knowledge**

Table 2 shows the results of the Pearson’s correlation analysis between the general task performance and users’ general topic knowledge. Users’ entry general task topic knowledge was positively correlated with their total number of statements in the final reports (r(22)=.406, p<.05), but not with the total number of facts (r(22)=.196, p=.359). The results indicate that the more entry knowledge users had with the general task topics, the more total number of statements they wrote in their reports.

The evolving general task topic knowledge did not show significant correlation with either the total number of statements (r(22)=.211, p=.076) or the total number of facts (r(22)=.074, p=.537).

The exit general task topic knowledge did not show significant correlation with either the total number of statements (r(22)=.231, p=.278) or the total number of facts (r(22)=.080, p=.709).

![Correlation r (p)]

<table>
<thead>
<tr>
<th></th>
<th>Total number of statements (general task)</th>
<th>Total number of facts (general task)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry general task topic knowledge</td>
<td>.406 (.049) *</td>
<td>.196 (.359)</td>
</tr>
<tr>
<td>Evolving general task topic knowledge</td>
<td>.211 (.076)</td>
<td>.074 (.537)</td>
</tr>
<tr>
<td>Exit general task topic knowledge</td>
<td>.231 (.278)</td>
<td>-.080 (.709)</td>
</tr>
</tbody>
</table>

*significant @ the 0.05 level

**Table 2. Correlation between general task knowledge and general task performance**

**Session-task Performance across Tasks**

A comparison of users’ session-task performance in the two types of tasks is shown in Table 3. No differences were found between the two tasks in either the number of statements (Mann-Whitney U(71)=559.5, p=.554) or the number of facts (Mann-Whitney U(71)=772.5, p=.160). This finding shows the same pattern, i.e., no difference, as that of the general task performance between two task types.

<table>
<thead>
<tr>
<th></th>
<th>Number of statements Mean (SD) (session task)</th>
<th>Number of facts Mean (SD) (session task)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Task</td>
<td>20.53 (10.50)</td>
<td>26.39 (14.57)</td>
</tr>
<tr>
<td>Parallel Task</td>
<td>18.94 (9.07)</td>
<td>28.11 (9.79)</td>
</tr>
<tr>
<td>Mann-Whitney U (p)</td>
<td>559.5 (.554)</td>
<td>772.5 (.160)</td>
</tr>
</tbody>
</table>

**Table 3. Session-task performance in two tasks**

**Session-task Performance across Sessions**

As is shown in Table 4, although descriptively, users had a greater number of statements and more facts in their reports in later sessions than in earlier sessions, no statistically significant differences were found between the three sessions in either the number of statements (Kruskal-Wallis
H(70)=3.56, p=.169) or the number of facts (Kruskal-Wallis H(70)=0.08, p=.961).

**Session-task Performance and User Knowledge**

As Table 5 shows, the number of statements in session-task reports was positively correlated with session-task topic knowledge (r(70)=.249, p<.05). This means the more knowledge the users had with the session-task topics, the more statements they wrote in their session reports. This is similar to the findings of the relationships between general task performance and general topic knowledge.

No significant correlation was found between users’ entry general task knowledge and their session-task performance measured by either the number of statements (r(70)=.089, p=.457) or the number of facts (r(70)=.084, p=.453). Although users’ entry general topic knowledge was correlated with the total number of statements (as reported above), their evolving general topic knowledge did not appear to correlate with their performance in individual sessions. Likewise, no significant correlation was detected between users’ evolving general task topic knowledge and their session-task performance measured by either the number of statements (r(70)=.188, p=.113) or the number of facts (r(70)=.061, p=.610).

<table>
<thead>
<tr>
<th>Task type</th>
<th>Knowledge group</th>
<th>Task knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>F(p)</td>
<td>Task type</td>
<td>Knowledge group</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Total number of facts</td>
<td>.247 (.625)</td>
<td>.026 (.874)</td>
</tr>
<tr>
<td>Total number of statements</td>
<td>.250 (.622)</td>
<td>.449 (.510)</td>
</tr>
</tbody>
</table>

Table 5. GLM analysis results: the effects of task type and topic knowledge on task performance

As can be seen from Table 6, knowledge group and task type did not show either main or interaction effects on both measures of task performance. Figures 1 and 2 depict the relationship between users’ knowledge, task, and performance. Although there was not a significant interaction effect, a tendency was shown that users with different levels of knowledge had different performance patterns in different tasks.

For the total number of facts, higher knowledge users did better descriptively, although not statistically significantly, when working with the parallel task than the dependent task, but low knowledge users did not show much difference in either task. In the parallel task, higher knowledge users did descriptively better than lower knowledge users, but in the dependent task, the pattern was the opposite: lower knowledge users did descriptively better than higher knowledge users.
For the total number of statements, higher knowledge users did equally well in both tasks, but lower knowledge users did descriptively better in the dependent tasks than in the parallel task. In the parallel task, higher knowledge users did descriptively better than lower knowledge users, but in the dependent task, both groups of users did equally well.

<table>
<thead>
<tr>
<th>F(p)</th>
<th>Task type</th>
<th>Knowledge group</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of facts</td>
<td>.511 (.478)</td>
<td>.053 (.818)</td>
<td>.450 (.640)</td>
</tr>
<tr>
<td>Total number of statements</td>
<td>.373 (.544)</td>
<td>.668 (.417)</td>
<td>.539 (.586)</td>
</tr>
<tr>
<td>Writing time</td>
<td>.042 (.839)</td>
<td>1.92 (.171)</td>
<td>1.128 (.330)</td>
</tr>
<tr>
<td>Search time</td>
<td>3.997 (.050)</td>
<td>.088 (.768)</td>
<td>4.412 (.016)</td>
</tr>
</tbody>
</table>

Table 7. GLM: The effects of task type, topic knowledge, and task session

Relationships among Task Type, Topic Knowledge, Task Session, and Task Performance

Users’ task performance was further examined in different task sessions, as shown in Table 7 and Figures 3 and 4. Here the entry general task knowledge was used. Figure 3 shows that in both tasks, higher knowledge users had descriptively more facts in their reports in later sessions than in earlier sessions, but lower knowledge users had descriptively fewer facts in session 3 in the dependent task, while in the parallel task, they had almost the same number of facts in all 3 sessions.

Figure 4 shows that higher knowledge users had descriptively more statements in later sessions than in earlier sessions in the parallel task, but they had descriptively fewer statements in later sessions than in earlier sessions in the dependent task. Lower knowledge users had the opposite patterns. In the dependent task, they had descriptively more statements in later sessions than in earlier sessions, but in the parallel task, they had descriptively fewer statements in later sessions than in earlier sessions.

For searching time, the pattern was different, and even opposite in some factors, compared to the writing time. Both higher and lower knowledge users spent significantly (F(72, 3)=4.412, p<.05) shorter time in the later sessions.
than in earlier sessions, in both tasks. When comparing between the two tasks, both higher and lower knowledge users had descriptively longer searching time in the dependent task than in the parallel task.

Figure 5: Relationship between topic knowledge and writing time in 3 sessions in both tasks

Figure 6: Relationship between topic knowledge and searching time in 3 sessions in both tasks

DISCUSSION

Task Type, Task Session, and Task Performance

Results show that users’ information use task performance, assessed in this study by the number of statements and the number of facts in the reports, did not show significant differences in the two task types: the parallel and the dependent. This was true for both the general tasks and the session-tasks. Although task type has been found in previous studies to be a factor that affects users’ information search including behaviors, process, and usefulness judgments (e.g., Lin, 2001; Liu & Belkin, 2010), in the current research, it did not show an effect on users’ work task performance. One explanation could be that the session-tasks had the same levels of difficulty, as evidenced by the analysis illustrated by Table 9 (Kruskal-Wallis test $\chi^2(2) = .276, p=.871$).

User Knowledge and Task Performance

Users’ knowledge, operationalized in the current study as their levels of familiarity with task topics, was found to correlate with task performance in some instances. For general task performance, users’ entry general task topic knowledge was found to have a positive correlation with the total number of statements in users’ final reports. For the session-task performance, users’ session-task topic knowledge also showed a positive correlation with the number of statements in the session-task reports. These results are intuitively reasonable in that those who are more knowledgeable could reach better task performance. This can be explained by further exploration of users’ behaviors and task performance, as discussed in the next sub-section.

User Knowledge and Search/Writing Behaviors

Our results show that users’ knowledge of session-task topics was positively correlated with session-task performance, and users’ time spent on writing was also positively correlated with session-task performance. However, there was no correlation between users’ topic knowledge and their writing time. Users with higher levels of topic knowledge did not seem to need longer writing time in order to generate task outcome with better performance – their topic knowledge probably contributed to a higher writing efficiency with the assigned topics. It would be beneficial to explore in future research how systems can assist and help increase low knowledge users’ performance but not decrease that of high knowledge users.


types in the current study were in the same domain, had the same task outcome requirements, and had the same levels of difficulty (Table 8). They were different than the tasks designed in previous studies that usually had different topics, complexity or difficulty levels, and so on.

<table>
<thead>
<tr>
<th>Difficulty type</th>
<th>Mean (standard deviation)</th>
<th>Mann-Whitney U(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-task</td>
<td>2.83 (1.34)</td>
<td>2.58 (0.90)</td>
</tr>
<tr>
<td>Post-task</td>
<td>2.92 (1.31)</td>
<td>2.08 (0.90)</td>
</tr>
</tbody>
</table>

Table 8. The pre- and post-task difficulty of two tasks

Results also demonstrate that users’ session-task performance did not vary across task sessions. Although previous studies in the literature found that task session (or stage) affects users’ information search including behaviors, process, and usefulness judgments (e.g., Lin, 2001; Liu & Belkin, 2010), in the current research, it did not show an effect on users’ work task performance. One explanation could be that the session-tasks had the same levels of difficulty, as evidenced by the analysis illustrated by Table 9 (Kruskal-Wallis test $\chi^2(2) = .276, p=.871$).

<table>
<thead>
<tr>
<th>Session</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24</td>
<td>2.5</td>
<td>1.319</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>2.33</td>
<td>1.274</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>2.63</td>
<td>1.715</td>
</tr>
<tr>
<td>Total</td>
<td>72</td>
<td>2.49</td>
<td>1.434</td>
</tr>
</tbody>
</table>

Table 9. Session-task difficulty ratings in 3 sessions
**Task Type, Topic Knowledge and Task Performance**

Neither task type nor users’ entry topic knowledge were found to affect users’ report writing performance. However, there was a tendency, although not statistically significant, that higher knowledge users had more facts in the reports in the parallel task than in the dependent task, and lower knowledge users had more statements written in the dependent task than in the parallel task. Further examination on users’ performance in the three sessions in both tasks confirmed the above tendency pattern.

These patterns can possibly be explained by users’ writing and searching behaviors. For the amount of time spent on writing, although higher knowledge users did not show much difference in the two tasks, but lower knowledge users had longer writing time in the dependent task than in the parallel task. Regarding the amount of time spent on searching, higher knowledge users did not seem to differ in the two tasks, but lower knowledge users had longer searching time in the dependent task than in the parallel task. Results on both the search and the writing time showed that longer writing time in a session did not necessarily meant shorter searching time, for example, in the case of the lower knowledge users working with the dependent task. Nevertheless, this may have to do with users’ abilities of comprehending and using the retrieved information in their task products. The confirmation of this speculation needs further analysis on how participants used retrieved information in the task products, which will be done in future research.

Another explanation of the pattern of topic knowledge, task type and task performance could be that higher knowledge users knew enough about the topic and also what information they needed: the parallel task asked them to find different aspects of the topic and saved their time searching and comprehending the information sources; on the other hand, the dependent task asked them to find hierarchically related information, which instead prevented them from using the information more easily and directly as in the parallel task. Oppositely, the lower knowledge users did not know enough about the topic and what information would be desirable: the dependent task gave them the overview information and could have helped them to more easily learn and master the task topic (although they may still have needed more time to search than the higher knowledge users), and accordingly better used them in the task products; on the other hand, the parallel task did not give them the overview of the task topic and may have prevented them from a comprehensive mastering of the topic in general and accordingly effectively using the located information in the task products.

The results were not statistically significant, which may be due to the small sample size of the current experiment, and a future study with a larger sample size could further test these. Nevertheless, the tendency pattern seems to suggest that systems can provide queries and/or webpages leading to different structures of the multi-aspect tasks according to users’ knowledge levels in order to help improve users’ task performance. For instance, for higher knowledge users, the system can decompose a task into parallel-structured subtasks, and for lower knowledge users, the system can decompose a task into dependent-structured subtasks. Again, further research on participants’ use of retrieved information in their products such as interviewing with them may hopefully confirm these suppositions.

**Limitations, Implications, and Future Studies**

The study has some limitations. The sample size is relatively small, which could have been at least part of the reason for few statistically significant results. Most of the participants were female, meaning there may be some interaction effect with the task topic being car purchasing. Also, care should be taken in generalizing the findings to all multi-aspect information use tasks.

Nevertheless, the study has some implications for system design. Scholars have been proposing that IR system evaluation should consider comprehensively how the system supports task accomplishment or goal achievement and consider supporting information use besides information search (Belkin, Cole, & Liu, 2009; Vakkari & Huuskonen, 2012). Our results show that users’ knowledge could affect their information use task performance, and that task structure may also play a role in influencing task performance. IR system design can try to improve lower knowledge users’ task performance, while not decreasing that of high knowledge users. One suggestion would be decomposing multi-aspect tasks into parallel subtasks for high knowledge users and providing query suggestions based on the different aspects of the multi-aspect task topic; and decomposing tasks into dependent subtasks for low knowledge users and providing query suggestions, for example, that can provide increasingly in-depth overview information sources. Future studies will further explore the effects of these factors, design prototypes of this suggestion of task decomposition, and conduct user studies to evaluate the effectiveness of this strategy.

**CONCLUSIONS**

This study found that in multi-aspect tasks, task structure and task session did not affect users’ task performance as measured by the number of statements and number of facts in the reports. Users’ entry knowledge of the general task topics correlated with their general task performance, but their evolving and exit knowledge did not. User’s session topic knowledge correlated with their session task performance. Furthermore, although not statistically significant, a tendency was shown regarding task structure and users’ entry general task topic knowledge on users’ task performance. Specifically, higher knowledge users had better performance in the parallel tasks than in the dependent tasks, and lower knowledge users had better performance in the dependent tasks than in the parallel tasks. This may be explained by users’ searching and writing behaviors: the effectiveness of finding useful pages and time allocated to writing were found to correlate with
task performance. These results seem to suggest that systems can be designed to provide queries and/or webpages leading to different structures of the multi-aspect task according to users’ knowledge levels in order to help them increase their task performance. These findings help understand factors influencing information task performance. They also have implications on IR system design in providing personalized support to information task accomplishment besides searching.

ACKNOWLEDGMENTS
Our thanks to IMLS for sponsoring the research experiment under grant number LG#06-07-0105-05.

REFERENCES