Grannies, tanning beds, tattoos and NASCAR: Evaluation of search tasks with varying levels of cognitive complexity

Wan-Ching Wu, Diane Kelly, Ashlee Edwards, Jaime Arguello

School of Information and Library Science

University of North Carolina at Chapel Hill

Chapel Hill, NC, 27599-3360 USA

[wanchinw, diane.kelly, aedwards, jarguello]@unc.edu

ABSTRACT

One of the most challenging aspects of designing an interactive information retrieval (IIR) study is the development of search tasks. In this paper, we present preliminary results of a study designed to evaluate a set of search tasks that were developed for use in IIR studies. We created 20 search tasks using five levels of cognitive complexity and four domains, and conducted a laboratory evaluation of these tasks with 48 undergraduate subjects. We describe preliminary results from an analysis of data from 24 subjects for 10 search tasks. Initial results show that, in general, as cognitive complexity increased, subjects issued more queries, clicked on more search results, viewed more URLs and took more time to complete the task. Subjects' expected and experienced difficulty ratings of tasks generally increased as cognitive complexity increased with some exceptions. When subjects were asked to rank tasks according to difficulty and engagement, tasks with higher cognitive complexity were rated as more difficult than tasks with lower cognitive complexity, but not necessarily as more engaging. These preliminary results suggest that behaviors and ratings are fairly consistent with the differences one might expect among the search tasks and provide initial evidence of the usefulness of these tasks in IIR studies.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval---search process; H.5.2 [**Information Interfaces and Presentation (I.7)**]: User Interfaces (D.2.2, H.1.2, I.3.6) ---evaluation/methodology

General Terms

Performance, Design, Human Factors

Keywords

Interactive information retrieval, search tasks, cognitive complexity, search behavior

1. INTRODUCTION

Search tasks are one of the most important components of interactive information retrieval (IIR) studies. In most studies, researchers assign search tasks to users in order to study search behavior and evaluate systems. In some cases, search tasks are ancillary to the study purposes but are needed for users to exercise systems, while in other cases search tasks are a key part of the study purpose and act as independent variables.

Search tasks can be quite complex and vary along a number of dimensions [6]; their development can be difficult and time consuming, and often requires specialized knowledge and skills. Development is complicated by the abundance of research demonstrating how variations in search tasks and search task properties can impact behavior [c.f., 4, 7, 8]. Unfortunately, there is little formal guidance about how to construct and evaluate search tasks [for exceptions see 3, 10] and for practical reasons (e.g., space) empirical reports usually do not provide a thorough description of how search tasks were generated or a full description of the tasks. Despite long-standing calls in the IIR and HCI communities for the development of standardized task sets, reference tasks and sharable tasks, little effort has been made to address these calls [7, 9, 10].

In this paper, we present preliminary results of a study designed to evaluate a set of search tasks that were developed using the cognitive complexity dimension of Anderson and Krathwohl's Taxonomy of Learning [1] a well-known education resource. This taxonomy and our tasks are described in more detail in the next section. The purpose of this study is to evaluate the tasks by collecting a series of data, both behavioral and self-report that allow us to understand and demonstrate differences among tasks. The ultimate goal of this work is to release these tasks to the research community. We also hope the framework will help others systematically develop their own search tasks.

2. SEARCH TASKS

In this study, we focus on the *cognitive process* dimension of Anderson and Krathwohl's Taxonomy of Learning (Table 1). Six types of cognitive processes are identified: *remember*, *understand*, *apply*, *analyze*, *evaluate* and *create*, with each type requiring increasing amounts of cognition and effort. While this taxonomy is traditionally used to create educational materials such as exercises and exam questions, we used it to construct search tasks similar to Jansen *et al.* [5].

Table 1. Anderson and Krathwohl's Taxonomy of Learning Objectives (Cognitive Process Dimension) [1]

Dimension	Definition
Remember	Retrieving, recognizing, and recalling relevant knowledge from long-term memory.
Understand	Constructing meaning from oral, written, and graphic messages through interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining.
Apply	Carrying out or using a procedure through executing, or implementing.
Analyze	Breaking material into constituent parts, determining how the parts relate to one another and to an overall structure or purpose through differentiating, organizing, and attributing.
Evaluate	Making judgments based on criteria and standards through checking and critiquing.
Create	Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure through generating, planning, or producing.

We selected four domains to use when creating the tasks: health, commerce, entertainment, and science and technology, and situated the tasks within scenarios geared toward our target subjects [3]. We used all the cognitive processes except *apply* because we were unable to create search tasks for this category that were distinct from the other categories. In total, we created 20 tasks (one for each cognitive process/domain combination). The tasks were initially created and used in another research project about vertical search [2]. The initial development consisted of pilot tests with six subjects and then a full user study with 28 subjects with tasks representing the first three levels of cognitive complexity and five domains. In this previous study, we found a significant relationship between cognitive complexity and interaction: as complexity increased, so too did queries issued, URLs visited, search result clicked and time spent conducting the search. These results encouraged us to continue development of the tasks. The tasks were updated and revised based on the findings from the previous study and subjected to additional critique and analysis. The 10 tasks used in the initial analyses reported in this paper are presented in Appendix A.

3. METHOD

A laboratory study with 48 undergraduate subjects was conducted to evaluate the search tasks. Data from 24 of these subjects is reported here. Each subject completed five search tasks from one domain and within each domain task types were rotated using a Latin-square. For practical reasons, we were unable to randomly assign subjects to domain. Subjects searched the open Web using Google and were asked to create responses to each task by typing answers and/or copying/pasting evidence that helped them arrive at their solutions. No time limits were given. Subjects completed Pre- and Post-Task Questionnaires and an Exit Questionnaire.

3.1 Pre-Task Questionnaire

The Pre-Task Questionnaire was divided into three parts: interests and knowledge, task complexity [4] and task difficulty. In this paper, we focus on the latter part (Table 2), which is related to the amount of effort subjects expected to expend completing various activities related to the search tasks, including results evaluation and integration, and determining when they had enough information to stop searching.

Table 2. Task difficulty items from Pre-Task Questionnaire

Item	Response
How difficult do you think it will be to *search* for information for this task using a search engine?	Not at all difficult Slightly difficult Somewhat difficult Moderately difficult Very difficult
How difficult do you think it will be to *understand* the information the search engine finds?	Not at all difficult Slightly difficult Somewhat difficult Moderately difficult Very difficult
How difficult do you think it will be to *decide* if the information the search engine finds is *useful* for completing the task?	Not at all difficult Slightly difficult Somewhat difficult Moderately difficult Very difficult
How difficult do you think it will be to *integrate* the information the search engine finds?	Not at all difficult Slightly difficult Somewhat difficult Moderately difficult Very difficult
How difficult do you think it will be to determine *when you have enough* information to finish the task?	Not at all difficult Slightly difficult Somewhat difficult Moderately difficult Very difficult

3.2 Post-Task Questionnaire

The Post-Task Questionnaire was divided into four parts. The first part contained three items that asked subjects to describe how they felt while completing the task, including their level of enjoyment, engagement and their abilities to concentrate. The second part consisted of two items that asked subjects to indicate the extent to which their interests in and knowledge of the topic represented by the task changed. The third part consisted of the five difficulty items presented in Table 2, with minor editorial changes to reflect past tense. The fourth part consisted of three items designed to elicit summative judgments from subjects. Again, because we have just recently finished this study, we focus on subjects' responses to the difficulty items.

3.3 Exit Questionnaire

The Exit Questionnaire asked subjects to rank order the tasks according to difficulty and engagement, and provided them with a free-text box to explain their rankings.

3.4 Participants

Subjects were recruited from the undergraduate student population at our university and consisted of 17 women and 7 men. Subjects' average age was 20 years old. The frequency of majors was: 10 sciences, 4 social sciences, 3 humanities, 6 professional schools and 1 undecided. Most subjects reported conducting information searches daily and having been searching for approximately 7-9 years.

4. RESULTS

4.1 Search Interaction

Search interaction included time spent on task (Time), number of queries issued (Queries), number of results clicked (Clicks on search engine results page: SERP), and number of URLs visited, including both SERP results clicked and links in web pages accessed. Results in Table 3 show a tendency for subjects to spend more time, issue more queries, click on more search results, and visit more URLs during tasks with greater cognitive complexity. A repeated measures ANOVA showed that all differences were statistically significant. Results from Bonferroni post-hoc tests showed that on all four measures Remember and Understand tasks differed significantly from Create tasks. For time and clicks on SERP, Remember tasks also differed significantly from Understand tasks.

Table 3. Cognitive Complexity and Interaction (Mean, Std)

	Time (sec)	**Queries**	**Clicks on SERP**	**URLs**
Remember	161.96 (167.43)	1.58 (.78)	2.17 (1.41)	3.33 (5.13)
Understand	344.29 (189.33)	1.71 (.99)	3.33 (1.71)	4.41 (2.54)
Analyze	496.33 (300.68)	2.42 (1.35)	3.46 (2.38)	4.58 (3.44)
Evaluate	545.83 (315.10)	2.33 (1.58)	3.92 (2.26)	5.54 (3.46)
Create	556.42 (341.20)	3.12 (1.92)	5.46 (3.66)	11.21 (9.34)
F(4, 92), *p*-value	18.22, p<.001	7.05, p<.001	8.72, p<.001	13.32, P<.001
Bonferroni post-hoc tests	R<U<A, E, C	R, U<C; R<A	R<U<C; R<E	R, U, A<C

4.2 Expected and Experienced Task Difficulty

We asked subjects how difficult they expected it would be to search for information (Search), to understand found information (Understand Info), to decide the usefulness of information (Decide), to integrate found information (Integrate) and to determine when found information was enough (Stop), and we asked the same questions after the actual search. Responses were coded such that 1=not at all difficult and 5=very difficult.

Results show that all responses ranged from 1.08-2.54, indicating that none of the tasks were perceived to be too difficult both before and after search. As cognitive complexity increased, in general, both pre- and post- values on all measures increased accordingly except for stopping (Figure 1). While it was likely easy to determine when enough information was obtained for Remember tasks, our participants found the other four tasks equally difficult to determine when to stop.

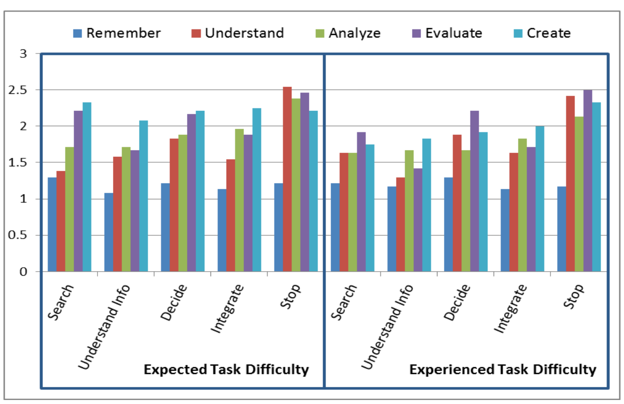

Figure 1. Expected and Experienced Task Difficulty

Results from an ANOVA show that cognitive complexity level had a significant effect on Expected and Experienced Task Difficulty for all five measures except the Search measure of Expected Task Difficulty (Table 4). Results from Bonferroni's post-hoc tests can also be found in Table 4.

	Expected*	**Post-hoc**	**Experienced***	**Post-hoc**
Search	6.26, p<.001	R, U<E, C	1.64 p=.169	-
Understand Info	4.04 p=.004	R<C	2.67 p=.036	-
Decide	5.07 p=.001	R<E, C	2.83 p=.028	R<E
Integrate	4.99 p=.001	R<A, C	2.62 p=.038	R<C
Stop	6.24 p<.001	R<U, A, E, C	6.54 p=.001	R<U, A, E, C

Table 4. Effect of cognitive complexity on Expected and Experienced Task Difficulty [*F (4, 115)]

4.3 Difficulty and Engagement Rankings

After completing all searches, subjects ranked tasks from 1 to 5 according to the overall difficulty (1=Least difficult; 5=Most difficult) and engagement (1=Least engaging; 5=Most engaging). As shown in Figure 2, the mean rank of each task type corresponded to the cognitive complexity level. However, in terms of engagement with tasks, subjects felt the most engaged when they searched for Analyze tasks and were the least engaged when they searched for Understand tasks.

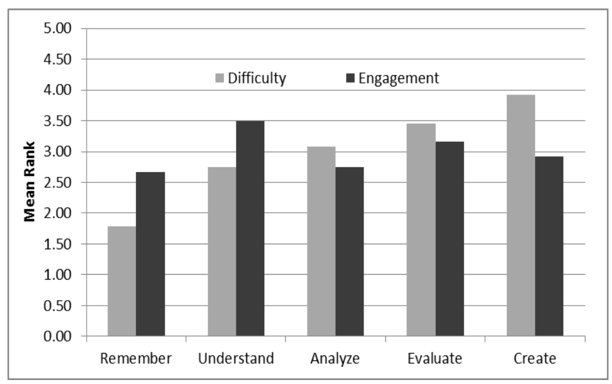

Figure 2. Mean Ranks by Difficulty and Engagement

5. CONCLUSION

The purpose of this research project is to develop and evaluate a set of search tasks that can be used in IIR studies. We created and evaluated ten search tasks from two domains representing five levels of cognitive complexity. We found that search interaction increased as the level of cognitive complexity increased, as well as most of the self-reported expected and experienced difficulty ratings. Moreover, subjects' rankings of task difficulty reflected the designated levels of cognitive complexity. However, the ranking of task engagement did not seem to be closely correlated with level of cognitive complexity. In future work, the full set of measures we collected will be analyzed as well as the results from the two domains not reported in this poster. These additional results will likely give us better insight into how subjects experienced these search tasks and the potential impact of domain, topic knowledge and interest. These additional results will also provide IIR researchers with better information for determining if, and how, they can use these tasks in their own research. In addition, we believe the framework itself will also be of benefit to those interested in designing their own search tasks.

In this study, we focused on one type of subject that is often represented in IIR studies, undergraduate students. Future studies might investigate these tasks with different populations of subjects to see if the general relationships hold, especially with regard to search interaction. In our initial work with these tasks, we studied people from our community and found similar results with respect to the interaction data (subjects did not complete the full battery of questions asked of our current subjects) [2].

IIR is a research specialty that spans international boundaries. We recognize that our tasks have some cultural biases built-in; this was necessary in order to create tasks that we thought would appeal to our target subjects. This is one barrier to creating search tasks that can be shared with an international research community. However, we believe many of our tasks have international appeal and can be modified for geographical relevance.

6. ACKNOWLEDGEMENTS

We thank Professor Barbara Wildemuth for her helpful advice and guidance.

7. REFERENCES

[1] Anderson, L. W. & Krathwohl, D. A. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

[2] Arguello, J., Wu, W.C., Kelly, D., & Edwards, A. (2012). Task complexity, vertical display and user interaction in aggregated search. *Proceedings of SIGIR '12*, Portland, OR.

[3] Borlund, P. (2003). The IIR evaluation model: A framework for the evaluation of interactive information retrieval systems. *Information Research, 8*(3), paper 152.

[4] Byström, K. & Järvelin, K. (1995). Task complexity affects information seeking and use. *IP&M, 31*(2), 191–213.

[5] Jansen, B. J., Booth, D., & Smith, D. (2009). Using the taxonomy of cognitive learning to model online searching. *IP&M, 45*, 643–663.

[6] Li, Y. & Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *IP&M, 44*, 1822–1837.

[7] Toms, E. G. (2011). Task-based information searching and retrieval. In I. Ruthven & D. Kelly (Eds.) *Interactive Information Seeking, Behaviour and Retrieval*. Facet Publishing, 43–59.

[8] Vakkari, P. (2003). Task-based information searching. *ARIST, 37*, 413–464.

[9] Whittaker, S., Terveen, L., & Nardi, B. (2000). Let's stop pushing the envelope and start addressing it: a reference task agenda for HCI. *Human Computer Interaction,* 15, 75-106.

[10] Wildemuth, B.M. & Freund, L. (2009) Search tasks and their role in studies of search behaviors. Paper presented at the *Third Annual Workshop on HCIR*, Washington D.C

Appendix A: Search Tasks

	Health	Science & Technology
Remember	You recently watched a documentary about people living with HIV in the United States. You thought the disease was nearly eradicated, and are now curious to know more about the prevalence of the disease. Specifically, how many people in the US are currently living with HIV?	You recently watched a show on the Discovery Channel, about fish that can live so deep in the ocean that they're in darkness most or all of the time. This made you more curious about the deepest point in the ocean. What is the name of the deepest point in the ocean?
Understand	Your nephew is considering trying out for a football team. Most of your relatives are supportive of the idea, but you think the sport is dangerous and are worried about the potential health risks. Specifically, what are some long-term health risks faced by football players?	You recently became acquainted with one of the farmers at the local farmer's market. One day, over lunch, he was on a rant about how people are ruining the soil. He was clearly upset, so you're interested in finding out more. What are some human activities that degrade soil fertility?
Analyze	Having heard some of the recent reports on risks of natural tanning, it seems like a better idea to sport an artificial tan this summer. What are some of the different types of artificial tanning methods? What are the health risks associated with each method?	You recently became involved with a conservation group that picks-up trash from local waterways. One of the group members told you that your work was important because it helps keep pollution out of the ocean. What are some of the different types of ocean pollutants? What environmental risks are associated with each pollutant?
Evaluate	One of your siblings got a spur of the moment tattoo and now regrets it. What are the current available methods for tattoo removal, and how effective are they? Which method do you think is best? Why?	You have noticed that online services such as Facebook have replaced face-to-face interactions. You can see the advantages of this style of communication, but your sibling argues that people are losing their ability to communicate face-to-face. In general, does use of computers for communication have a positive or negative impact on people's face-to-face social skills?
Create	Your great granny's doctor has told her that getting more exercise will increase her fitness and help her avoid injuries. Your great granny does not use the Internet and has asked you to create an exercise program for her. She is 90-years old. Put together two thirty-minute low-impact exercise programs that she could alternate between during the week.	After the NASCAR season opened this year, your niece became really interested in soapbox derby racing. Since her parents are both really busy, you've agreed to help her build a car so that she can enter a local race. The first step is to figure out how to build a car. Identify some basic designs that you might use and create a basic plan for constructing the car.