

# Complex Search Task: How to Make a Phone Safe for a Child

Sophie Rutter

School of Information Science  
University of Sheffield  
Sheffield, UK  
sarutter1@sheffield.ac.uk

Verena Blinzler

University of Regensburg  
Regensburg, Germany  
verena.blinzler@stud.uni-regensburg.de

Chaoyu Ye

Mixed Reality Lab  
University of Nottingham  
Nottingham, UK  
psxycy1@nottingham.ac.uk

Michael B. Twidale

Graduate School of Library and Information Science  
University of Illinois at Urbana-Champaign  
Champaign, IL, USA  
twidale@illinois.edu

Max L. Wilson

Mixed Reality Lab  
University of Nottingham  
Nottingham, UK  
max.wilson@nottingham.ac.uk

## ABSTRACT

There are many factors in task design that might make it ‘complex’: having multiple components, having multiple cross-dependent components, tasks that involve comparison, evaluation, estimation, or learning. In this paper, we discuss a case study of a complex task we may consider to be highly natural, a common concern for many people, and one that ‘should’ have a clear answer, but doesn’t: how do you make a phone safe for a child. For this question, there is a lot of opinion online, many possibilities for actions, many variations in hardware and software, but ultimately no one clear and correct answer for everyday phone users. We found very little objective behaviours that separated people in terms of performance but instead have begun to identify some successful tactics that are not directly linked to domain knowledge.

## KEYWORDS

Information Seeking, Complex Tasks, Search Literacy

## 1 INTRODUCTION

Designing complex tasks for a user study is hard. Wildemuth and Freund [10] synthesized all the aspects of search tasks that might make them exploratory in nature, which include: learning goals, general topics, open-ended topics, multi-focus needs, multi-faceted needs, uncertain aims, ill-structured problems, and which are “not too easy”. Exploratory tasks will therefore involve long and dynamic searching processes, which are accompanied by other information and cognitive activities, such as analysing, organizing, and decision making [10]. Choosing complex tasks to embody some portion of these factors for user studies, and indeed comparing the observed behaviours with those seen in other user studies, is a non-trivial process. As a community, we may have studied many complex tasks, but we are perhaps still a

long way from a comprehensive and discriminatory model of how people solve different complex tasks, and thus how search systems can support them.

It is also hard to generate complex search tasks, especially those being performed on uncontrolled collections like the World Wide Web, as much online content is user generated content that reflects their attempts at solving possibly similar questions. This means that it is difficult to use previously designed complex tasks, because the solutions often become available online<sup>1</sup>; indeed, such pages could appear during the data collection period of a study.

This paper presents a case study of a task used in our ongoing Search Literacy project, where our final task was the product of many failed attempts at selecting a task that would engage participants for more than ten minutes, without asking them to ignore certain websites or informational pages. We asked participants to find out how to make a phone safe for a child (Figure 1) and below we review how this task fits into the exploratory task facets identified by Wildemuth and Freund [10]. We then review our initial findings as a means to discuss approaches to evaluating performance in such complex tasks.

## 2 STUDY AND TASK CONTEXT

The aim of our research was to study *Search Literacy*, and how it affects searchers; we wanted to examine and compare how competent searchers attempted a task, in comparison to less competent searchers. A person with good search literacy should be able to resolve e.g. technology problems when they are out of their depth in the domain. The secondary ongoing aim, therefore, is to find design recommendations that would help a person to become more competent.

---

CHIIR 2017 Workshop on Supporting Complex Search Tasks, Oslo, Norway.  
Copyright for the individual papers remains with the authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors. Published on CEUR-WS, Volume 1798, <http://ceur-ws.org/Vol-1798/>

---

<sup>1</sup> Although interesting new complex search tasks are posted on Dan Russell’s Search Research blog, many people often post their solutions.

A friend of yours has recently bought a new phone (the one provided here). Sometimes their child uses the phone. Your friend has asked for your help.

1) They do not want to unnecessarily restrict their child from downloading Apps. They do, however, want to ensure that **they do not hear more than mild bad language and they do not see any violence directed towards humans**. Is this possible? What should they do?

2) They would like to set up a **separate profile** for their child but have been unable to do so. Why would your friend find this difficult? Can you do this for them? If not, why not?

3) What else would you **recommend** your friend do to make the phone safe for a child to use?

Please help your friend by searching the Internet (on the laptop provided) to find solutions. When you have found a solution (s) you should implement these on the phone provided.

**Figure 1: Task Description**

## 2.1 Task Choice

Our chosen task, shown in Figure 1, was to find out how to make an android phone safe for a child. The goal of the task was primarily to **learn** about how best to make the phone safe (although to simulate a real work task [4], they were asked to make changes to an actual Samsung phone running the Lollipop version of android’s operation system). The task was **general** in terms of the topic, and the task had multiple targets. The task was also **multi-faceted**, in that there were three related sub-tasks that could be achieved. One of the sub-tasks was open-ended in that there was no one correct answer to find. Two of the sub-tasks were more specific in that participants needed to find a solution to a problem and make direct recommendations. A key aspect of the overall task was **uncertainty**. Firstly, in that the best way to make a phone safe for a child is highly discussed online, as is general internet safety for children and the views are conflicting. Secondly, for the second sub-task much of the information available online (at the time of the study anyway) was misleading. As such this task was also **ill-formed**, in that it was based upon what a person might want to achieve, rather than what can be achieved. The task was also successfully **dynamic** and **long**, in that the majority of participants used all 20 minutes without completing all three parts of the task. The task, therefore, was also certainly “**not too easy**”, especially in that creating profiles in the second aspect of the task was not achievable in that version of android on that make of phone. Finally, based upon Wildemuth and Freund’s factors [10], the task involved many related information and cognitive activities, including sensemaking, comparing, and decision making.

Several versions of the task, as well as alternatives, were trialled in pilot studies. Many alternatives that we tried, including e.g. how to set up a Chrome browser, eventually all

had instructional videos. A key factor in the success of our final task is that it is a) debated fundamentally in terms of child protection approaches, b) achieved in many ways (protection vs prevention, etc.) and c) implemented differently for different hardware and software versions. Crucially this meant that ready solutions were not available.

## 2.2 Participants and Protocol

We recruited 39 participants using two strategies. Initially, we recruited participants to: take part in a study about solving technical problems using a search engine. We then used a self-assessment scale for search literacy and technical competence, based upon the EU Digital Competence framework [5]. After determining that our initial sample of participants had mostly high search literacy and high tech domain knowledge, we later recruited people with posters asking ‘do you ask **other people** to solve your **tech problems**?’. Consequently, we aimed for a mix of participants: 17 had high search literacy and high domain knowledge (HH), 13 had high search literacy and low domain knowledge (HL), and 9 had low search literacy and low domain knowledge (LL). Because of the chosen domain, however, we did not have any participants that we classified as low search literacy and high domain knowledge (LH), implying that having good “tech knowledge” came along with higher search literacy in our sample. We also later classified people as being successful at different performance levels, described below, and gathered information about other domains of knowledge, including parenting and experience with different mobile phone platforms.

After gathering informed consent, participants were presented with the task in the form of a simulated work task [4] and given 20 minutes to make progress on it. We did not allocate specific time periods to each sub-task, and so participants could work towards the larger task by attempting the subtasks in any order, or indeed in combination; finding advice for part 1 often meant encountering information for part 3, for example. The screen of the phone and the laptop were both recorded, and the movement between them was recorded using a GoPro Hero 4 camera. The interaction with the Chrome browser was also comprehensively logged using a custom extension<sup>2</sup>. After the time was up, participants completed a short questionnaire, before reviewing their laptop screen recording as part of a post-task interview. This post-task interview allowed us to capture a reflective cued-retrospective think aloud [9] of their search processes and gain insight into their cognitive activities. The browser and phone was reset between participants, to remove revisitation indicators for subsequent participants, however the study was performed in a Computer Science department and so the results could have been affected by our location.

The study was approved by the school’s ethics board, and participants received a £10 Amazon Voucher as remuneration for their time. Although the task was not entirely achievable, no participants exhibited signs of distress at being unable to

<sup>2</sup> <https://github.com/kelvinve/ChromeExtensionForWebData>

complete the task. In fact, most participants were enthusiastically engaged such that they did not want to stop searching after the allotted time. In fact, most believed that child safety was so important that the information *should* be clearly available, and if anything were frustrated that it was not.

### 3 INITIAL RESULTS AND DISCUSSION

We broke participants' performance into three levels, based on a point-rating given to all three sub-tasks. The three groups are typically (but not exclusively) characterized by their resolution of the second part: 1) those that were unable to find basic information, including the location of the phone's settings (N=9), 2) those that thought they had completed the second part, but had an incorrect solution (N=21), 3) those that correctly concluded that part two wasn't possible (N=9). Participants in the top group also tended to make more than one recommendation in part three of the task.

#### 3.1 Objective Behavioural Differences

We examined many metrics of search behaviour, from number of queries and page views, to average query length, speed of interactions, and dwell time. We found very few differences between participants, when broken down by both performance and by self-assessed search literacy and domain knowledge. In fact, much of the time-based data was affected by the participant interacting with the phone as they testing the found information. Long periods of dwell time were not because participants were reading results, but testing them. Indeed, longer dwell times associated with good searching techniques [8] were often more evident in participants that struggled with the task. Two further activities are considered to be good searching techniques: evaluating search results and thus deeper clicks in the SERP. However, we saw that the most effective participants clicked very quickly on top results only, without examining the source. Whilst our initial results indicate that low-performing participants clicked deeper in the search results. In interviews, high performing participants indicated that they simply trusted the search engine to put reputable results on the top, but judged the utility of the result after clicking on them. We plan to release the logged behaviour data as part of a dataset in the future.

#### 3.2 Search Process and Tactical Differences

Overall, the majority of differences that we saw between high and low performing participants was more to do with their search process and use of different tactics. Based on Bates' search tactics [2,3] and Barry & Schamber's relevance criteria [1], we qualitatively analysed the post-task interviews to evaluate the tactics that participants used to solve the task. We found that participants used tactics to (1) manage the task, the tactics that are used to answer the tasks and manage the search process, (2) control the search, the moves made to direct what information is received and to manage information across multiple devices, and (3) evaluate and use information, the tactics that participants use to select objects. For each of these areas of concern, participants had tactics that they could use in

isolation or in combination to progress the search towards the resolution of task problems.

Although we are still finishing this analysis, early results indicate that there are tactics associated with domain knowledge and tactics associated with good performance, and that the two are not an exact match. Different tactics, for example, are available depending on the domain knowledge of the participant. For example, when selecting search results, those with more tech knowledge evaluated the date field in the snippet because they were aware that information about technology quickly dates. However, this tactic did not necessarily improve performance. An example of a tactic that is associated with good performance, rather than domain knowledge, was narrowing the query early in the process. By including information about the phone (e.g. model, make etc.) the results returned were more specific to the task. This tactic was used by high performers, including those with high and low domain knowledge.

### 4 CONCLUSIONS

In this study, we set users a very complex search task, involving a general problem that was multi-faceted, and where the solutions were not easily recognizable. Participants had to engage in information and cognitive activities during the task, as well as interacting with and testing solutions on a physical phone in between searching. Overall, we found that objective log data was not the best source for evaluating the open-ended, dynamic, and extended periods of searching involved in resolving complex search tasks. Instead, we were able to evaluate the searching from the tactics that participants employed. Use of different tactics, made the largest difference in task performance. We conclude that striving to convert logged behaviour into tactics (e.g. [6]) is important future work for evaluating complex search tasks. Further, we expect that future search user interfaces should a) encourage participants to move between more and less specific searches when important, and b) help searchers to identify key concepts in results and perform secondary searches about them.

### REFERENCES

- [1] Barry, C.L. & Schamber, L., 1998. Users' criteria for relevance evaluation: a cross-situational comparison. *IP&M*, 34(2-3), pp.219-236.
- [2] Bates, M.J., 1979. Idea tactics. *JASIST*, 30(5), pp.280-289.
- [3] Bates, M.J., 1979. Information search tactics. *JASIST*, 30(4), pp.205-214.
- [4] Borlund, P. & Ingwersen, P., 1997. The development of a method for the evaluation of interactive information retrieval systems. *JDOC*, 53(3), 225-50.
- [5] Ferrari, A., 2013. DIGCOMP: A framework for developing and understanding digital competence in Europe.
- [6] He, J., Qvarfordt, P., Halvey, M. and Golovchinsky, G., 2016. Beyond actions: Exploring the discovery of tactics from user logs. *IP&M* 52(6), pp.1200-1226.
- [7] Laxman, K., 2010. A conceptual framework mapping the application of information search strategies to well and ill-structured problem solving. *Computers & Education*, 55(2), pp.513-526.
- [8] Vakkari, P., Luoma, A. & Pöntinen, J., 2014. Books' interest grading and dwell time in metadata in selecting fiction. In *Proc. IIX'14*, (28-37). ACM.
- [9] Van Gog, T., Paas, F., Van Merriënboer, J.J. & Witte, P., 2005. Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, 11(4), p.237.
- [10] Wildemuth, B.M. & Freund, L., 2012. Assigning search tasks designed to elicit exploratory search behaviors. In *Proc. HCIR '12*, Article 4.