

# The Myth of Find: User Behaviour and Attitudes Towards the Basic Search Feature

Fernando Loizides  
Future Interaction Technologies Lab  
Computer Science Department  
Swansea University  
csfernando@swan.ac.uk

George Buchanan  
Future Interaction Technologies Lab  
Computer Science Department  
Swansea University  
g.r.buchanan@swan.ac.uk

## ABSTRACT

The ubiquitous within-document text search feature (*Ctrl-F*) is considered by users to be a key advantage in electronic information seeking [1]. However what people *say* they do and what they *actually* do are not always consistent. It is necessary to understand, acknowledge and identify the cause of this inconsistency. We must identify the physical and cognitive factors to develop better methods and tools, assisting with the search process. This paper discusses the limitations and myths of *Ctrl-f* in information seeking. A prototype system for within-document search is introduced. Three user studies portray shared behaviour and attitudes, common between participants regarding within-document searching.

## Categories and Subject Descriptors

H.0 [Information Systems]: General; H.m [Information Systems]: Miscellaneous

## General Terms

Experimentation, Human Factors

## Keywords

Information Triage, Digital Libraries, Information Seeking, Within-Document Searching

## 1. INTRODUCTION

Searching for information can be a tedious, time consuming and laborious task. Depending on the information required and the repository used by the information seeker, the information can sometimes seem unobtainable, in that it is taking more time to locate than is worth or available.

Most document readers, for example PDF (Portable Document Format) readers, have the commonly known ‘Find’ function, otherwise known as *Ctrl-f* (see fig.1). This, in its basic form allows a user to insert a search query, that the program matches to identical text in the document and takes

the user to the first point of occurrence. The program then continues to go through all the matches of the query in a linear fashion, presenting the results with query word highlighting to the user.

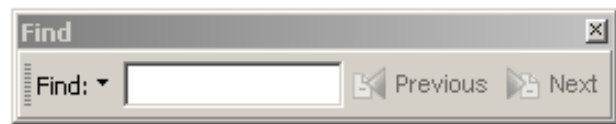


Figure 1: Basic Within-Document Search Feature

The basic search function in document reader software provides very little functionality: only a literal match facility. Users who do not know exactly what they are looking for may provide a query that does not match all related parts of the document relevant to their need. Naive users may not appreciate the lack of intelligence in this technology.

It is not surprising then, that when performing document triage (a process by which a user decides on the relevance of documents to his or her information need) users often only look at the first match when using the *Ctrl-f* feature [1]. This resembles the findings in [7], namely that most information seekers will only look at two result list pages or less when querying a search engine. Buchanan and Loizides also report on how users frequently choose not to use the search function although it is regarded as a ‘key advantage’ that electronic reading and searching has to offer [1]. One hypothesis is that although users regard the feature as useful overall, they realise the limited potential and the tedious process of linearly going through exact search terms and use it selectively. This results in the users choosing to simply skim documents rather than guessing query terms to match.

## 2. RELATED WORK

There is limited research on effective electronic tools for searching within documents. [2] and [3] compare two within-document search tools (ProfileSkim and FindSkim). FindSkim highlights words in the document that match those from the user’s query or variants of these. ProfileSkim builds on the features of FindSkim. A document is divided into ‘fixed length sections’ called ‘tiles’. Using a method called *relevance profiling* a relevance score is calculated for every *tile*. The higher the score of the tile, the more relevance it is expected to have to the information seeker’s need. An analogue interactive navigational bar allows the user to overview

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

JCDL’08, June 16–20, 2008, Pittsburgh, Pennsylvania, USA.  
Copyright 2008 ACM 978-1-59593-998-2/08/06 ...\$5.00.

the results for each tile in the document and navigate accordingly within the document.

The first of the three user studies presented in this paper reports data from a user study presented in [1]. The other two studies uncover further details of the use of the find feature. The first user study describes the observation of participants while performing document triage under three conditions; using paper documents, an electronic results list and an electronic folder.

User behaviour in the studies coincides with Gary Marchionini's model for information seeking [6]. The actions seen span from the *Problem Definition* stage and ranges to the *Reflection and Stop* stage. The focus of this paper though in terms of analysis, concentrates on the *Query Formulation*, *Results Examination* and *Information Extraction* stages.

### 3. USER STUDY DESCRIPTION

The findings described in this paper originate from the results taken from three separate user studies; all three being different in nature, in order to remove any bias and explore different areas. These include: a diary study; an observational video recorded study; an expert review.

#### 3.1 User Study 1: Overview of Document Triage

The full description of this user study can be found in [1]. 30 participants were selected for the user study. The aim of the user study was to explore how users triage information (in document format) between different mediums; electronic and physical.

#### 3.2 User Study 2: Diary Study

Eight participants were asked to perform a diary study of their search habits in terms of triage of information. The users spanned across a wide range of disciplines; law, computer science, physics, engineering, psychology, criminology, media and finance. An initial semi-structured interview was conducted in which the participants were asked to verbally explain how they go about a typical search. Direct questions probed to what kind of material they search for in general, frequency of searches and methods used for searching. We also elicited their perceptions of the advantages and disadvantages of physical and electronic searching means. If the topic of *Ctrl-f* and its usefulness did not surface from the participants' comments, a direct question was made to elicit details.

The participants recorded their actions, their thoughts and feelings when performing triage in their own time. The searches could be either work related or of personal interest. Participants were given between one week and one month to complete the diary in order for the results to be as representative as possible and least obtrusive. An intermediate interview took place when the participants were ready to hand in their first few diary entries. During this interview the existing data was examined and questions were asked to better understand the searches and behaviours while they were still fresh in the participants' mind. A final interview was conducted at the end of the study to obtain direct feedback from the participants. During each interview we took care not to overemphasize the importance of *Ctrl-f* and thus affect the study. In total there were 35 searches documented by the participants.

### 3.3 User Study 3: Expert Review

The third user study consisted of a group of eight experts in the field of computer science. The experts were presented directly with the theory that *Ctrl-f* is substantially limited in its current state and interviewed for their thoughts and opinions. A probe of software was presented with more advanced search facilities, both for evaluation and to provoke response to the proposed theory. Although a detailed description of the alternative feature is beyond the scope of this paper, it is sufficient to say that it separates the current document into sections (simulated paragraphs) and treats each section as a unique document. An integrated algorithm processes the query in relation to each section and produces a ranking for each section based on a *Term Frequency and Inverse Document Frequency* calculation. The technique of stemming [8] is also applied for improved results. Highlighting focuses the words in the document sections that match to the query terms. At the time of the user study the system was in early development stages in terms of aesthetics and limited to specific file formats, but fully operational for the users.

The participants were permitted to ask questions relating to the system at any point. While the experts were engaged in the system, notes were taken on their actions to be later scrutinised and analysed for more information. After the users had finished answering the tasks (and performing exploratory interaction), they were asked a series of questions to discern their thoughts in terms of query search tactics used and their view on the limitations of the current generic search feature after interacting with a more 'capable' system.

## 4. FINDINGS

In this section we present the behaviours and actions observed from the user studies. We identify reasons for these in terms of cognition. Furthermore, limitations or features not present in *Ctrl-f* are identified. The usefulness of these non-existent features are justified by the findings.

### 4.1 The Use of the Search Feature

41 of 46 participants from all three studies considered the search feature to be a 'key advantage' in electronic searching. Several claim to use this feature regularly. The popularity of the search feature in practice is not consistent with the subjective claims made for it by the users.

The findings in [1] report that from the 20 users taking part in the user study only four used *Ctrl-f*. From 243 documents read in full, only 11 resulted in the use of the feature. Even when utilised, use of *Ctrl-f* was limited, with only three of those occasions moving past the first result.

During the initial interview before commencing the diary study every participant commented that the search feature played a very important part in seeking for information electronically. This claim was only made after the interviewer specifically asked the participants their views on the feature and not from the initial comments on the benefits of electronic searching.

Interestingly, despite the value given to *Ctrl-f*, none of the participants in the diary study recorded its use. When reviewing results in the final interview, two participants claimed to have used it but did not remember when. One stated that he may have used it and five of the participants answered that they did not make use of the feature, thus undermining the claim of usefulness. This contradicts the

claim of usefulness for the search feature. One participant commented that he had ‘never actually used *Ctrl-f* before’. The behaviour holds for short searches as well as longer time wise searches, with times ranging from one minute quick searching, to three hour intense triage of information<sup>1</sup>.

## 4.2 Analysing User Behaviour

### 4.2.1 Knowing What To Look For

In our first triage study users were observed not to linearly traverse more than the first search result. We asked for a detailed explanation. The responses included ‘I guess I don’t really know what I’m looking for’ and ‘My search was too specific and I could tell from the first result that it wouldn’t be what I was looking for’. The papers in the first user study were related to touch screens. Participants attempted to search for relevant document sections using synonyms to ‘touch screens’ such as *touch screen* and *interface*. Since the papers chosen for the study were at least partly relevant (as would often be the case with search results), the search words would often appear. This did not provide the participants with enough indication of the importance of the paper in relation to the search queries.

One of the expert participants from the third user study, commented that he usually does not type a whole word when using *Ctrl-f*: e.g he used the search query term ‘*hallu*’ instead of ‘*hallucinogenic*’. He explained that ‘The less letters I have, the most probability I have to get the spelling correct’. It is no wonder then that users put a great deal of trust in searching using search engines (as we will see later) which are able to suggest words close to an incorrect spelling that the user might write. This is a feature in which the generic search feature lacks.

### 4.2.2 Search Engine Reliance

The diary study participants showed an overwhelming reliance on using the internet to search for information. All 30 electronic searches performed by the diary study participants used the World Wide Web. 28 of these electronic searches, utilised a search engine as the starting point: e.g. one participant wished to search for cheap tickets using the world wide web. Instead of directly going to the airline’s website (which was known to him from the start), searched for the airline’s name on a search engine (which was also his home page) and clicked on the web site from the results list.

This behaviour manifests itself more deeply into the cognitive thinking of individuals, whose reliance in search engines increases for performing tasks other than the ones they were designed for. ‘Users are just so ingenious at doing what you didn’t expect them to do’ [4]. This is evident from the behaviour of one diary study participant who claimed that he searches solely using electronic means and mainly on-line search engines to web search rather than *Ctrl-f*. His reasoning was that ‘you can even find how to boil an egg on-line nowadays, so there is no need to search anywhere else’. His diary evidence justified his claims. He used the on line search engine in a way which made it a customised within-document search facility. Instead of using the engine to locate a web page to browse for information he relied on the abstract provided by the engine result list, in order

to give him the information required for short searches. If the information was not prominent in the abstract he would rarely open the page.

Furthermore, when more information was required, the participant would open a web page, and if at first glance (including a small scroll) did not locate the area of the web page which he desired, he would return to the engine’s result list and choose another result or change his query. He commented ‘I am confident it (the search engine) will find me a page where I can just see the result quickly’. This behaviour is similar to the behaviour of the participants of the first user study who were given a results list to triage. The few who decided to open the full document did not spend time reading to depth. The first page was the most important part for the users who seldom scrolled to skim the rest of the document.

### 4.2.3 Perceived Versus Actual Behaviour

There is a consistent discrepancy between the behaviours users report in interview and data from direct observation. In our diary studies there is also a gap between recorded actions and strategies reported in interview. Although the participants’ behaviour gives evidence of a lack of engagement with the system there is also another factor, more cognitive in nature that is evident from their comments. When presented with the prototype system four of the participants mentioned that it would be a good idea to make the feature ‘look closer to *Ctrl-f*’. The participants have a sense of familiarity and confidence in the system which, although not extensively used, still rates strongly in familiarity.

## 4.3 Absent and Inadequate Features

### 4.3.1 Limited Scope of Search

One limitation clearly evident in the *Ctrl-f* feature is the scope of its search. As mentioned, *Ctrl-f* only detects exact phrase matches within a document. The fact that the participants were not looking for minor details (in the case of the first user study participants) rendered the search practically useless to the triage process. The feature’s inability to detect content and context relevance, hinders a seeker’s decision, consumes time and conceals material causing users to dismiss valuable material.

One participant does not use the *Ctrl-f* feature until the second half of the documents is reached. When asked, the participant commented that the feature was only used as a last resort, after other search methods were performed in order to verify a decision rather than to make it. Another, uses the feature for the first four documents and after realising its limited effectiveness, triages the remainder without the help of *Ctrl-f*.

It was mentioned earlier that users find the search feature limited in that if an exact phrase match is not found then results that are in fact relevant, can be bypassed. One user tried to go around this by using less letters for a search term, thus minimising the margin of error for spelling. But spelling is not the only danger that may mislead the information seeker into missing relevant information. Participants often missed words on account of the tense or way in which words were used. For example the word ‘burn’ would not be detected in a search with ‘burning’ or ‘burned’. The technique of stemming [5] used in information retrieval to eliminate such problems and could be integrated into the

<sup>1</sup>This data is based on 30 of the 35 searches, which were made electronically by the diary study participants. The others used physical search methods

search feature (as it has been in the prototype system) to improve searches.

Stemming is only one of the techniques used in order to improve searching within this context. A very highly regarded method often used by on-line search engines is that of *query expansion*. This has been highly researched and refined since its origins [5]. *Automatic query term expansion* produces terms to be used in addition to the original query which can assist in the locating of relevant information if the corpus (in our case the document) does not include the search queries typed in by the user. Using morphology, synonyms and correction of spelling, recall can be improved.

### 4.3.2 Viewing Area

Four experts expressed a preference to having the search procedure and results contained within an area separate to the documents, rather than seeing the highlighted results on the document itself. Three preferred to have the option of both; to be able to switch between the two depending on the task solely at hand. Only one participant preferred displaying results on the actual documents.

### 4.3.3 Use with a small screen device

Mobile devices are increasingly becoming more ubiquitous, widespread and essential in everyday life [4]. Although the scope of this specific research was not focused on small screen devices, the participants were probed for an initial insight on how the *Ctrl-f* feature could be influential in such a domain. One specific issue raised with our experts was reading and searching on small screen devices; namely what information did the expert participants (if any) read on small screen devices and how did they search for information on their small screen device. Four participants mentioned web browsing using their mobile phones while the other four only mentioned text messages. None of the participants mentioned reading documents using a small screen device and were very unreceptive of the suggestion that it was feasible to perform such a task.

Seven of the participants commented that they would be willing to use the search feature on a small screen device to search through messages and web pages or ‘perhaps reference books quickly’ on the condition that there existed an easier way to input search terms. The two dominating suggestions were a larger keyboard and via voice recognition. From these suggestions it is evident that the lack of searching use of mobile devices lies partly with physical means and hardware. New, more intuitive and user friendly methods (such as voice recognition) is more important on mobile devices than the accuracy of results at this time.

## 5. FUTURE WORK

In this paper we have reported shortcomings in the basic search function of document reader software. We have analysed user interaction behaviour and identified the reasons behind these. From our findings, it is evident that there is a need for improved within-document searching tools. An experimental within-document search facility was introduced, with more advanced features to *Ctrl-f*. The capabilities of this feature currently facilitate some of the user needs recognised.

It is our aim to develop this software in order to target further user requirements by automating the already existing behaviours. We aim to use this software to also further investigate information seeking and document triage.

## 6. CONCLUSIONS

It is evident that there is a discrepancy between what people say they do and what they actually do. Three different user studies with participants from a wide range of levels and disciplines helped expose behaviours and actions. Users regard *Ctrl-f* as an advantage of electronic information seeking, but rarely choose to employ the feature when searching for information. The few cases in which it is used exposes the limitations that the feature has.

Other tools such as on-line search engines are used by the information seeker to work around the shortcomings of *Ctrl-f*. There is currently little research on and development of within-document search tools. As we have seen there is an evident need to do so. The prototype presented in this paper, and currently under development, aims at progressing the technology available to the information seeker and advance knowledge in information seeking techniques and methodologies.

## 7. ACKNOWLEDGEMENTS

To our eight experts for devoting their time and sharing their views.

## 8. REFERENCES

- [1] G. Buchanan and F. Loizides. Investigating document triage on paper and electronic media. In *Proceedings of the European Conference on Research and Advanced Technology for Digital Libraries*, number 35, pages 416–427, 2007.
- [2] D. J. Harper, I. Koychev, Y. Sun, and I. Pirie. Within-document retrieval: A user-centred evaluation of relevance profiling. *Inf. Retr.*, 7(3-4):265–290, 2004.
- [3] D. J. Harper, I. Koychev, and Y. Sun. Query-based document skimming: A user-centred evaluation of relevance profiling. *European Conference on Inf. Retr.*, pages 377–392, 2003.
- [4] M. Jones and G. Marsden. *Mobile Interaction Design*. Wiley, 2006.
- [5] S. J. K. and J. D.M. The use of automatically obtained keyword classifications for information retrieval. *Information Processing and Management*, (3):175–201, 1970.
- [6] G. Marchionini. *Information Seeking in Electronic Environments*. Cambridge University Press, 1995.
- [7] A. Spink, B. J. Jansen, D. Wolfram, and T. Saracevic. From e-sex to e-commerce: Web search changes. *Computer*, 35(3):107–109, 2002.
- [8] C. van Rijsbergen, S. Robertson, and M. Porter. *New models in Probabilistic Information Retrieval (British Library Research and Development Report, no. 5587)*. Cambridge, 1980.