

# A Longitudinal Study of Exploratory and Keyword Search

Max L. Wilson, mc schraefel  
School of Electronics and Computer Science  
University of Southampton, UK  
{mlw05r, mc}@ecs.soton.ac.uk

## ABSTRACT

Digital libraries are concerned with improving the access to collections to make their service more effective and valuable to users. In this paper, we present the results of a four-week longitudinal study investigating the use of both exploratory and keyword forms of search within an online video archive, where both forms of search were available concurrently in a single user interface. While we expected early use to be more exploratory and subsequent use to be directed, over the whole period there was a balance of exploratory and keyword searches and they were often used together. Further, to support the notion that facets support exploration, there were more than five times as many facet clicks than more complex forms of keyword search (boolean and advanced). From these results, we can conclude that there is real value in investing in exploratory search support, which was shown to be both popular and useful for extended use of the system.

## Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Information filtering, Query formulation, Search Process. H.3.7 [Digital Libraries] User issues. H.5.2 [User Interfaces]: Evaluation/methodology, Interaction Styles.

## General Terms

Measurement, Design, Human Factors, Verification.

## Keywords

Faceted, Search, Keyword, mSpace, Browsing, Longitudinal, User, Study.

## 1. INTRODUCTION

Digital Libraries are looking for strategies to better support users in gaining access to their collections so that the service is more effective and valuable for its users, especially in competitive commercial settings. Despite previous investigations into new forms of search [1, 11], we still know little about how they are realistically used and subsequently whether their value outweighs the added challenges to provide them [7]. With an aim to understand the real value that different exploratory and keyword search features provide to users, we present a longitudinal study of a digital library interface called mSpace [6] (Figure 1). The study ran for a month to get past common obstacles of user studies, such as new feature novelty or familiarity.

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, or 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 2004 ACM 1-58113-000-0/00/0004...\$5.00.

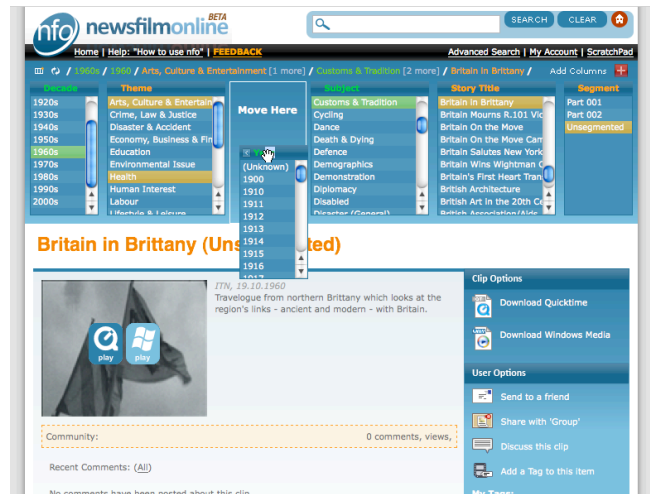


Figure 1: mSpace browser on the Newsfilm Online collection

In the following sections, we first discuss the different styles of faceted browsing as a function to support exploratory search. We then briefly describe mSpace to provide context to the study. In Section 3 we describe the collection and methods used in the longitudinal study and present the results from both regular communication with known participants and log analysis of a larger group of regular users. Finally, Section 4 concludes with the contributions of the research to digital library interface development and describes how the investigation will be continued for a longer period of time.

## 2. RELATED WORK

Here we present related work on mSpace and search techniques that support access to digital library collections.

### 2.1 Search and Faceted Browsing

Aside from advances in keyword searching, such as Boolean search [2] and interactive query refinements [4], other approaches have examined adding classifications to collections to enhance the ways that users can refine and constrain their queries. Hierarchical categories have been used successfully for many years to break down collections into groups and sub-groups that can be browsed and used to narrow down the range of documents relevant to a search. For unclassified collections, clustering can be used to automatically identify similar attributes of multiple documents within a collection [11]. In 2006, Hearst suggested that carefully designed facets of meta-data provide better support than automatically generated clusters [3].

Research into faceted browsing styles has produced two main approaches: traditional faceted browsing and directional column-faceted browsing; Wilson *et al.* have examined the differences between three faceted browsers that cover both styles [10]. In the

traditional style of faceted browsing, selecting an item in a facet universally filters each of the facets and the collection, so that selections may be made in any facet to further refine the results. In directional column-faceted browsers, including iTunes<sup>1</sup>, facets are usually presented horizontally in a line and filter only to the right. This allows for more inter-facet relationships to be shown, such as all the Artists in a Genre and all the Albums by one of the Artists. These two styles are described further by Wilson *et al.* [9].

Typically, novel approaches to support search are to carry out task-oriented user studies, where users are given an interface and task to complete. Research into the Open Video Project investigated the speed of video previews with such a user study [8]. Much of the research into mSpace has also been designed in this way [5, 9]. In 2007 Capra *et al* carried out a study of faceted browsers, which specifically included learning tasks to investigate the features designed to support exploratory search [1]. The study, however, was unable to show the specific benefits of exploratory features. Below, we present a longitudinal study into how exploratory and keyword forms of search are realistically used in a digital library service.

## 2.2 mSpace

mSpace, shown in Figure 1, is an example of a directional column-faceted browser that provides the added information about the relationships within an annotated digital library collection. A full breakdown of its functionality is given by schraefel *et al* [6], but we provide a brief overview below to give context to the study and results.

mSpace provides both facets and a strong keyword search function, including boolean queries, an advanced search form, and interactive query refinements. Aside from the default columns laid out from left to right across the top of the screen, columns can also be moved or deleted and other optional columns can be added. This gives the user control over the column-by-column relationships revealed. Columns to the right of a selection are then filtered, including a goal facet that lists all the documents; a separate typical search result list is shown below the columns too. Related items in columns to the left of a selection, which would remain after filtering if they were to the right, are highlighted to guide the user to refinements upstream [9]. Users may also select multiple items in a column, which acts as an OR union, and may filter individual columns to match specific string sequences.

To support decision-making, mSpace provides examples of documents, which can be shown on demand for any item in any facet. These are chosen at random, unless any specific metric is given, and have been shown to support users in their search [5]. This feature, however, was not included for the video library described below for funding and bandwidth reasons. mSpace also provides the ability to keep, share, and discuss items with social tagging, comments, and groups. Overall, mSpace and its gestalt of features has been shown to support more types of users and in more ways than other implementations of faceted browsing [10].

## 3. LONGITUDINAL STUDY

Here we present the collection and study methods used during a longitudinal study of exploratory and keyword searching, followed by the results found. Our hypothesis was that users

would exhibit exploratory search behavior during early use of the system and then move to directed keyword search in return visits.

### 3.1 Apparatus

Recently, mSpace has been applied to the Newsfilm Online Archive provided by ITN Source<sup>2</sup>. The collection contains around 3,000 hours (around 60,000 individual clips) of ITN and Reuters news footage taken from the 1900s to the current date. Each item has been annotated by the British Universities Film and Video Council<sup>3</sup> (BUFVC) with 25 facets of meta-data that include: a 3 level hierarchy of theme, a 4 level hierarchy of time, and facets such as series, country and language. Multiple formats of video, PDFs of original bulletin scripts, and still images for every second of each clip are also included in the available multimedia. As part of the development, the service is being evaluated in terms of its support for education and users.

### 3.2 Method

In participating over the month-long period, each participant was required to spend a few hours a week minimum using the service for their own needs or interests. Further, participants were required to respond to a weekly email containing both optional suggestions for the following weeks use and questions about the previous week's experiences. No training was given, so as not to disturb natural first-time experience, but participants were directed to the website help files if needed.

**Data Collection.** During the four-week period, user activity with the site was logged against IP addresses. An online forum was also provided for participants to discuss their experiences in three areas: feature requests to gain an idea of where users may have felt limited; problems with current features to gain an idea of where users struggled to use the current features; and interesting experiences with the service to gain insight into otherwise unmonitored use of the system. The longitudinal study was also started and concluded by telephone interviews with each of the known participants.

**Consent.** Each participant was informed of the nature and frequency of participation that was expected over the four-week period and explicitly accepted these terms through an online form. Along side this agreement, each participant submitted demographic details that provided context to the comments submitted during the study.

### 3.3 Participants

The study was made up of 11 known participants with varying experience, in terms of both news media and online digital libraries, and varying interest in the forthcoming service. 7 men and 4 women took part. Employment ranged from PhD student, to professor, to librarian. A further 11 anonymous users were logged using the system, but not involved in qualitative communications.

### 3.4 Results

Below we present the results of the study, in terms of overall feature use, typical interaction patterns, and use of facets.

---

<sup>1</sup> <http://www.apple.com/itunes/> - Apple – iPod + iTunes

---

<sup>2</sup> <http://www.itnsource.com/> - Home Page – ITN Source

<sup>3</sup> <http://www.bufvc.ac.uk/> - British Universities Film & Video Council

### 3.4.1 Overall use of features

The overall use of the different features found within the mSpace interface is shown in Table 1. The table breaks down keyword search into 3 forms, where Boolean and Advanced represent more complex forms of search. Considering the three types of keyword search provided, there is a very even use of keyword searching and column-facets in the system. By user and visit, however, we can see a tendency for keywords to be involved in more sessions, but columns used more often within sessions.

Columns can also be used to make more expressive and rich search queries, and we can see that there are more than five times as many column clicks as there are advanced forms of keyword searching. In all but one visit that using columns, two or more columns were used together. We can see from the breakdown by visit that many more sessions involved complicated queries produced by the columns than by boolean or advanced search.

When asked about the general good aspects of the site, one user stated: "I like being able to scan for rich content... I like the fact that I can stay at a high level while searching i.e. see the column search structure... [The column structure] helps place [the clips] in a broader context... I like having that clearly visible." This comment represents many, which revealed that sometimes column use was passive and more for learning about the collection.

**Table 1: Table showing the total, per user, and per visit breakdown of mSpace features used.**

\* Not including the Story Title column.

\*\* In the advanced search fields.

Feature	Total	Per User avg (n)	Per Visit avg (n)
Keyword Searches	211	10.55 (20)	6.03 (35)
Boolean Searches	30	2.50 (12)	2.31 (13)
Advanced Searches	15	2.50 (6)	2.14 (7)
Column Clicks*	252	18.00 (14)	7.88 (31)
Used/Viewed Tags	24	4.00 (6)	2.40 (10)
Viewed Other Users	5	1.25 (4)	1.25 (4)
Used/Viewed Groups	9	2.25 (4)	2.25 (4)
Managed Account	40	4.00 (10)	2.11 (19)
Auto-complete **	26	3.71 (7)	2.60 (10)
Long Column Paging	62	6.89 (9)	4.13 (15)
Extra Results	61	5.08 (12)	4.69 (13)

Table 1 also includes usage for more social functions within mSpace. We can see that although users managed their account frequently and used the tagging functionality, there was little interaction with other users and groups. Three possible reasons have been identified for this disjoint of behavior. First, it is possible that participants used the tagging for their own benefit and were not interested in other users. Second, that the system had not reached a critical mass of users to make the social aspects worthwhile. Third, that the participants found it hard to find other users and groups. Our qualitative communications reveal that the third of these options was the case for most of the participants. When asked about other users of the site, one participant stated: "I haven't explored that functionality yet, but when I have looked, I have yet to see other peoples searches or tags". Further discussion

and other comments revealed that the increased behavior on tagging and managing profiles was in an effort to be present in the social network, but the lack of interaction with other users and groups shows that participants found it hard to find them. This provides some direction for future work.

The final items in the column tell us about a) how often the auto-complete function was used in the advanced search, b) how often long lists were paged to cope with data transfer, and c) how many times users viewed more than the first page of results or reordered them. No specific trends were found from these fields, but the last shows that there is some demand for manipulating result sets and the amount of auto-complete used indicates that there were an average of two fields used per advanced search. The relatively small number of column paging suggests that most users did use the facet-columns to narrow their search results.

### 3.4.2 Typical Interaction Patterns

A summary of search sessions, separated into four categories depending on the use of columns and keyword searching, is shown in Table 2. Column use ignores the Story Title facet, as unlike the other facets, or keyword searching, it is a visualization of results and so does not provide any filtering. As well as showing the total number of sessions that exhibited each behavior, the table includes a breakdown of the first and second visits; 3<sup>rd</sup> or later visits were too infrequent to reveal any patterns and are not presented at this time.

Keyword Search Only and Column Use Only represent clear types of patterns that exclusively use one of the two methods of finding clips from the collection. The Keyword Search First and Column Use First conditions represent user patterns where one method was used first and the alternative method was used to refine or continue narrowing down results. Only 47 of the 55 total sessions involved either the columns or the keyword search. The remaining sessions included users who scrolled and clicked directly on the Story Title column, viewed previously found clips from their account or found clips from community tags.

**Table 2: Table showing the behavior patterns of column and keyword use each session.**

Behavior Pattern of Column and Keyword Search Use	Visit		Total
	1	2	
Keyword Search Only	10	1	16
Keyword Search First	5	3	10
Columns Used First	3	4	9
Columns Used Only	3	3	12

Overall we can see a fairly balanced use of both keyword searches and columns, with 35 involving columns and 31 involving facets. The breakdown of the first two sessions, however, provides an insight into the flow of use. Almost half of first visits only involved keyword searching. 85% of the first visits used keyword searches in some way, and only 50% used columns. Only 15% use columns exclusively. In the second visits, 91% used the columns as part of their search, with only 9% using exclusively keyword searches and 25% using columns alone.

In support of these findings, one participant commented that she had understood how the columns were reacting to her discoveries by the end of the first session, and that she felt confident about how to use them. Another participant stated: "Working with this

interface has already suggested new directions and avenues to explore in relation to the clips. So (in my case) it certainly helps knowledge building and lateral thinking.”

### 3.4.3 Use of different facets

The spread of columns, or facets, used during the month-long period, is shown in Figure 2. We can see an expected difference in use between the columns that are presented automatically to the user, and the facets that the users can add as desired. In support of the chosen facets, one user stated that the default columns were the most appropriate, of the available columns, to her needs for finding clips to use in her classes.

We can see also that the thematic facets were used more often than the temporal facets. This is present in both the default and optional facets, where Theme and Subject were used more than Decade and Year, and Subtopic was used almost double the amount of times that Month was involved. The two popular additions that were not thematic or temporal were Series Title (within the hierarchy of clips) and Country, which is geographic.

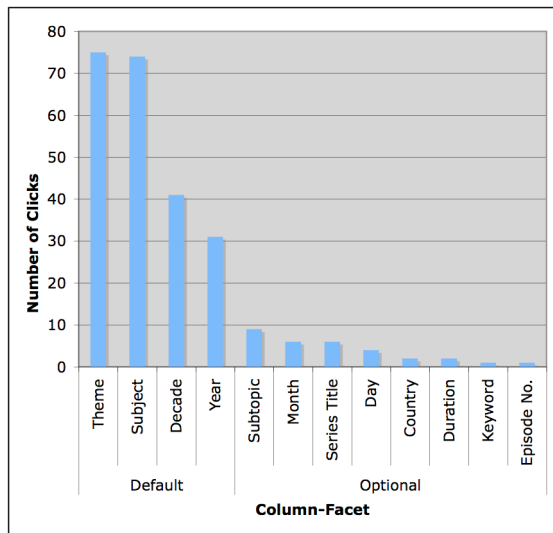


Figure 2: Breakdown of the column-facets used over the four week period, where default are displayed automatically and optional can be added explicitly by the user.

## 4. DISCUSSION AND CONCLUSIONS

The aim of the study above was to investigate the extended and real-life use of exploratory and keyword styles of search. We expected that early use would be exploratory and subsequent use would be made up of keyword searches. In contrast, however, we have seen that, given an initial hesitation to use exploratory features, extended use of the system was made up equally of both styles. Our results show that the column facets were used as often as keyword searches throughout the study period, and that the facets were used both passively to understand the structure of the collection and actively to produce more expressive queries, as multiple columns were nearly always used together.

Our contributions above are: a) carrying out a month-long user study with 11 known and 11 anonymous participants over a significantly large collection; b) using the study to compare a variety of keyword and exploratory search features; and c) gaining insight into usage patterns over time. Our results have shown that not only are exploratory forms of search used as often as keyword

search, but that they were often used to produce more expressive queries. These results are contrary to some skepticism about the value of exploratory search features and surpassed our own hypothesis that exploratory features would be used in the early exposure to a system. Given these discoveries and the desire to improve access for digital library users, we can conclude that there is significant value overcoming the known challenges [7] and providing exploratory search features as part of a digital library interface.

## 5. ACKNOWLEDGMENTS

Thank you to the participants of the longitudinal study who gave regular time and feedback about the service. The collection is funded by JISC, provided by ITN Source, annotated by the BUFVC, and hosted by EDINA.

## 6. REFERENCES

1. Capra, R., Marchionini, G., Oh, J.S., Stutzman, F. and Zhang, Y. Effects of structure and interaction style on distinct search tasks. *Proceedings of the 2007 conference on Digital libraries*. 442-451.
2. Croft, W.B., Turtle, H.R. and Lewis, D.D. The use of phrases and structured queries in information retrieval. *Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval*. 32-45.
3. Hearst, M.A. Clustering versus faceted categories for information exploration. *Commun. ACM*, 49 (4). 59-61.
4. Ruthven, I. Re-examining the potential effectiveness of interactive query expansion. *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*. 213-220.
5. schraefel, m.c., Wilson, M.L. and Karam, M. Preview Cues: Enhancing Access to Multimedia Content, School of Electronics and Computer Science, University of Southampton, 2004.
6. schraefel, m.c., Wilson, M.L., Russell, A. and Smith, D.A. mSpace: improving information access to multimedia domains with multimodal exploratory search. *Communications of the ACM*, 49 (4). 47-49.
7. Smith, D.A., Owens, A., schraefel, m.c., Sinclair, P., André, P., Wilson, M.L., Russell, A., Martinez, K. and Lewis, P., Challenges in Supporting Faceted Semantic Browsing of Multimedia Collections. *The Second International Conference on Semantic and Digital Media Technologies (SAMT2007)*, (Genova, Italy, 2007).
8. Wildemuth, B.M., Marchionini, G., Yang, M., Geisler, G., Wilkens, T., Hughes, A. and Gruss, R. How fast is too fast?: evaluating fast forward surrogates for digital video. *Proceedings of the third ACM/IEEE-CS joint conference on Digital libraries*. 221-230.
9. Wilson, M.L., André, P., Smith, D.A. and schraefel, m.c. Spatial Consistency and Contextual Cues for Incidental Learning in Browser Design, School of Electronics and Computer Science, University of Southampton, 2007.
10. Wilson, M.L., schraefel, m.c. and White, R.W. Evaluating Advanced Search Interfaces using Established Information-Seeking Models. *JASIST*.
11. Zamir, O., Etzioni, O., Madani, O. and Karp, R.M. Fast and intuitive clustering of web documents. *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*. 287-290.