Keyword Search: Quite Exploratory Actually

Max L. Wilson
Future Interaction Technologies Lab
Swansea University, UK
m.l.wilson@swansea.ac.uk

ABSTRACT
This short position paper describes some evidence found that counters the argument that there are better ways to support exploratory search than keyword search. Instead, this paper suggests that keyword search actually provides people with the freedom to search in relation to their own current state of understanding, rather than in the terms controlled by a search system. The challenge for future exploratory search systems, therefore, may be to maintain and enhance such freedoms.

INTRODUCTION
Some of the main arguments for research into exploratory search are that there are times when keyword search is not sufficient to support users. Such occasions include times when users who are unsure about a certain domain of information, uncertain about the terminology used by a search system, or unsure, even, about their own information needs [7]. Alternatively, therefore, many have been trying to support users in more exploratory conditions with alternative visualizations and user interfaces. Faceted browsing, clusters, and tag clouds, for example, are techniques that are designed to expose the structure of, or relationships within information to users, so that they can better understand a domain of information.

So why is it that keyword search persists? In some occasions, as described below, users have even preferred keyword search during exploratory tasks. While this may be because of people's familiarity with keyword search, the argument being made here is that exploration involves activities for which keyword search can be quite appropriate. The core of these learning activities, for example, is in making sense of how unfamiliar information fits in with a user's current understanding. It is potentially important therefore, that exploration allows users to freely express their current understanding. Further, however, hypothesis testing is also an important aspect of sensemaking, where searchers, as they learn, may want to see how results change according to their own ideas and developing conclusions.

EVIDENCE FOR KEYWORD SEARCH
Above, it is suggested that there is some evidence for when users have preferred keyword search for more exploratory activities. In our own research, for example, we have seen that users found the facets in the mSpace browser useful more often for expressing multiple compound constraints in queries, than during exploration [10]. In another study, Capra et al. compared the RB++ browser and an un-configured Endeca[1] interface to the Bureau of Labor Statistics website[2]. First, the website, which of course has been designed for the dataset, performed well for all tasks. Further, however users specifically noted, during exploratory tasks, the lack of keyword search in the RB++ browser [3] (now included in the latest version).

More recently, our own research has created an analytical evaluation method [11] that can inspect search interface designs for how they support users in each of 16 searcher conditions. This method was used to evaluate, for example, the interfaces in the above examples [8]. Further, Google's keyword search was analysed, as shown in Figure 1, where the 16 profiles are described in Figure 2. These 16 search conditions range from users who know exactly what they want, and how to describe it (profile 16) to those who are learning and do not know what they will find (profile 1) [1].

In Figure 1, it might be noticed that the least supported searcher profile by keyword search is not profile 1, but profile 5, where users are scanning for an unknown document to take away, by recognizing it when they see it. This represents more browsing behaviour, where the user is trying to use keywords to describe a particular target that they are hoping exists. The support for exploration, however (towards profile 16) actually increases. Conversely, the most supported profiles are those where the user is trying to find a known target, by recognizing, and using keywords in their head. This process is actually better supported, with the help of query suggestions and spelling corrections, than users who know exactly what they want and can specify it, where users have to pick the terms that will most likely put their desired target at the top of the results list.

SENSEMAKING
Making sense of information revolves around a user bridging a gap between their own knowledge and new information they have found [4]. In analysing how people hand-off information from one person to another, during

[1] http://www.endeca.com

[2] http://www.bls.gov

shift changes for example, pitching information at the right level of knowledge and understanding for the receiver, is important [6]. During any sensemaking process, therefore, it should be important to see how their own state of knowledge, however superficial, affects results.

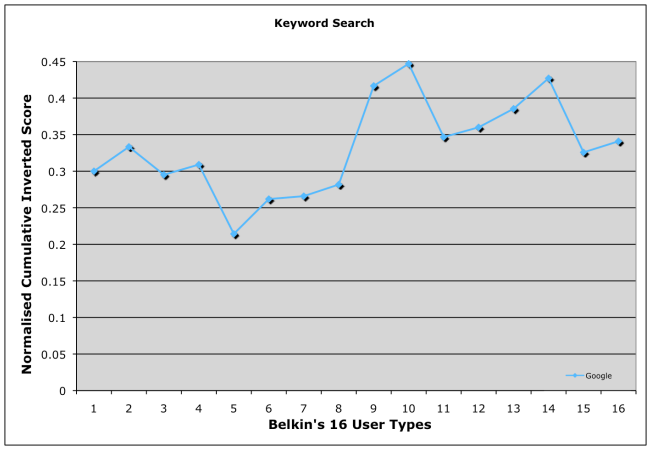

Figure 1: An analysis of keyword search across different searcher profiles, where 16 is the most knowledgeable about their target, and profile 1 represents those learning and exploring [8].

ISS	Method	Goal	Mode	Resource
1	Scan	Learn	Recognize	Information
2	Scan	Learn	Recognize	Meta-Information
3	Scan	Learn	Specify	Information
4	Scan	Learn	Specify	Meta-Information
5	Scan	Select	Recognize	Information
6	Scan	Select	Recognize	Meta-Information
7	Scan	Select	Specify	Information
8	Scan	Select	Specify	Meta-Information
9	Search	Learn	Recognize	Information
10	Search	Learn	Recognize	Meta-Information
11	Search	Learn	Specify	Information
12	Search	Learn	Specify	Meta-Information
13	Search	Select	Recognize	Information
14	Search	Select	Recognize	Meta-Information
15	Search	Select	Specify	Information
16	Search	Select	Specify	Meta-Information

Figure 2: The 16 searcher profiles from Belkin et al. [1].

People's own terms can also have a significant affect on memory and information processing. In a study of recalling blogs that participants had previously tagged, Budiu noted that participants performed best when they had tagged it using their own terms rather than the terms within the blog itself [2]. One possible hypothesis from these results is that users may perhaps struggle to interact with unfamiliar terminology laid out in faceted classifications, when they might rather try to communicate their own state of understanding. Even within facets of metadata, users are given the task of trying to find metadata they recognise, which, for all they know, may not be a valid option within the facet. At this point, it may be less effort for the user to say 'this is what I know', which is undoubtedly the way conversations would go when seeking the support of experts, or librarians as it used to be. There may be, in fact, no simpler way to express one's knowledge than to enter terms they understand into an empty box.

SO WHAT DOES THAT MEAN FOR HCIR?

The root of the argument being built here, is that free-text search, is so called because it gives people freedom. The challenge for exploratory search and HCIR, therefore, is to try and maintain or incorporate *freedom* into interface designs or new visualizations. With many HCIR interface features, like faceted browsing, involving classification schemes built from the data or constructed from the domain of information, this may be challenging. Clustering engines, for another example, cluster around the data or metadata, and cluster labels could mean nothing to the user at all. While it is not uncommon for facets to be filtered by keyword searches [5], or, as in mSpace, for highlights to appear in facets, which relate to a result found in a keyword search [9], it might be of more exploratory value to provide stemming and support for synonyms to highlight related terminology in facets.

Another challenge for HCIR design, based on what we know of sensemaking and handoffs, maybe to monitor users and then try to pitch information at their level. It might be that dynamic faceted systems, which select the appropriate facets to show at any one time rather than simply all possible facets, may meet this requirement to some extent already. It might also be possible, however, to modify the terminology in facets, or vary the language in result lists, to terms that the user would understand. Understanding users though, of course, is a hard challenge.

I by no means have the answers here, but the core of the challenge to the HCIR community will be to properly, beyond the hypothesis of a position paper, investigate the question: why is it that keyword search persists, and is often helpful for exploratory search? It will be this discovery that will allow us to try and replicate the benefits in future designs. Until then, however, the challenge is, while leveraging the benefits of metadata, to try making freedom the core of our human computer interaction designs for information retrieval.

CONCLUSIONS

The aim of this position paper has not been to suggest that the study of exploratory search is not important, or that research into alternative visualizations is not important. There are times, for example, especially where multiple or explicit constraints might be applied, such as in e-commerce, where faceted metadata is particularly useful. Instead, the aim of this position paper has been to highlight that there are elements of the keyword-response paradigm that are actually quite appropriate for exploratory search. While the challenge is to properly find out why keyword search has performed well in exploratory search, until then, the position here is that we should try to replicate keyword search's freedom in our future exploratory search designs.

REFERENCES

1. Belkin, N.J., Marchetti, P.G. and Cool, C., Braque: design of an interface to support user interaction in information retrieval. Information Processing and Management, *29*, 3 (1993). 325-344.
2. Budiu, R., Pirolli, P. and Hong, L., Remembrance of things tagged: how tagging effort affects tag production and human memory. In *CHI'09*, ACM New York, NY, USA (2009), 615-624.
3. Capra, R., Marchionini, G., Oh, J.S., Stutzman, F. and Zhang, Y., Effects of structure and interaction style on distinct search tasks. In *Proc. JCDL 2007*, ACM Press (2007), 442-451.
4. Dervin, B., Foreman-Wernet, L. and Lauterbach, E. Sense-Making Methology Reader: Selected Writings of Brenda Dervin. Hampton Press, 2003.
5. Hearst, M., Design Recommendations for Hierarchical Faceted Search Interfaces. In *Design Recommendations for Hierarchical Faceted Search Interfaces* (2006).
6. Sharma, N., Sensemaking Handoff: When and How? Proceedings of the American Society for Information Science and Technology, *45*, 1 (2009). 1-12.
7. White, R.W., Kules, B., Drucker, S.M. and schraefel, m.c., Introduction. Communications of the ACM, *49*, 4 (2006). 36-39.
8. Wilson, M.L. An Analytical Inspection Framework for Evaluating the Search Tactics and User Profiles Supported by Information Seeking Interfaces, University of Southampton, 2009, 249.
9. Wilson, M.L., André, P. and schraefel, m.c., Backward Highlighting: Enhancing Faceted Search. In *UIST'08*, ACM Press (2008), 235-238.
10. Wilson, M.L. and schraefel, m.c., A longitudinal study of exploratory and keyword search. In *A longitudinal study of exploratory and keyword search*, ACM Press (2008), 52-56.
11. Wilson, M.L., schraefel, m.c. and White, R.W., Evaluating Advanced Search Interfaces using Established Information-Seeking Models. Journal of the American Society for Information Science and Technology, *60*, 7 (2009). 1407-1422.