Slicing and dicing the information space using local contexts

Hideo Joho and Joemon M. Jose

Department of Computing Science
University of Glasgow
17 Lilybank Gardens
Glasgow G12 8QQ, UK

hideo,jj@dcs.gla.ac.uk

Abstract. In recent years there has been growing interest in faceted grouping of documents for Interactive Information Retrieval (IIR). It is suggested that faceted grouping can offer a flexible way of browsing a collection compared to clustering. However, the success of faceted grouping seems to rely on sufficient knowledge of collection structure. In this paper we propose an approach based on the local contexts of query terms, which is inspired by the interaction of faceted search and browsing. The use of local contexts is appealing since it requires less knowledge of the collection than existing approaches. A task-based user study was carried out to investigate the effectiveness of our interface in varied complexity. The results suggest that the local contexts can be exploited as the source of search result browsing in IIR, and that our interface appears to facilitate different aspects of search process over the task complexity. The implication of the evaluation methodology using high complexity tasks is also discussed.

1 Introduction

We use search engines to carry out a wide range of search tasks with various goals [1, 2]. Some tasks are easy, some are difficult. One of the elements that affects task difficulty is known to be task complexity [3]. Task complexity can be seen as a degree of uncertainty involved in a search task. Uncertainties such as what information is necessary for searching, how to find required information, or how to recognise required information are all part of task complexity [4]. As such, the perception of task complexity is subjective, thus, a high complexity task for one searcher might be a set of low complexity tasks for another. Studies have shown that task complexity is likely to affect how searchers interact with search interfaces [5–8]. For example, Marchionini [5] suggests that searchers tend to browse more frequently in a high complexity (HC) task while they tend to conduct a direct search in a low complexity (LC) task. This suggests that the user interactions that should be supported and facilitated by search interfaces should vary over task complexity. In particular, it has been pointed out that the study on the design and evaluation of search interface has been limited with regard to high complexity tasks [9].

In this paper, we propose a novel search interface inspired by a faceted approach to exploring search results. The proposed interface is designed to offer a greater control in the way search results are organised and explored. The design of the proposed interface was motivated by the following factors. Firstly, a set of keywords that co-occurred with query terms in the document surrogate were used to organise search results as opposed to clustering. While clustering techniques have been investigated as a means of structuring search results (e.g., [10, 11]), it often underestimates the diversity of topics discussed in a single document, thus, a document is categorised into a single cluster. Therefore, in a high complexity task where searchers are not necessarily aware of the dimensions of search topics and results, clustering appears to be less appropriate. On the other hand, "a faceted approach provides different ways to slice and dice the information space, which allows people to look at the information space from different perspectives" [12]. We were especially motivated by this type of flexible interaction achieved in the faceted approach.

Another factor is a nature of dimension extraction. In facet-base applications, the facets are often extracted in a static manner from the document collection or metadata [13]. Therefore, the extraction of facets is often a computationally expensive process. One also needs to be able to access the entire collection to create a comprehensive set of facets. In our interface, a set of keywords are automatically extracted from the local contexts of query terms, and used as the dimensions to slice and dice the information space. This not only allowed us to implement the facet-like interaction using existing search engines, but also made it easy for searchers to understand why these dimensions were presented to browse search results.

Finally, the proposed interface provides an independent area called a workspace in the interface. The importance of workspace in supporting information seeking processes has been suggested [9] and evaluated in an image retrieval [14]. In the proposed interface, the workspace area is provided in addition to the conventional search results area to facilitate user's exploratory activity.

The rest of paper is structured as follows. Section 2 provides the design and implementation of our search interface. Section 3 discusses the experimental methodology. Section 4 presents the results and analysis. Finally, Section 5 concludes the paper and discusses our future work.

2 Proposed search interface

Our search interface was implemented as follows. When a query was submitted to the interface, the query was sent to a back-end search engine (Google) with an additional option set to retrieve 100 results, as opposed to the default of 10. No further querying was performed to the search engine until another query was submitted to the interface. The interface parsed the returned result and displayed the first 10 retrieved records in the *Main Result* area (See Fig. 1 for a screen shot

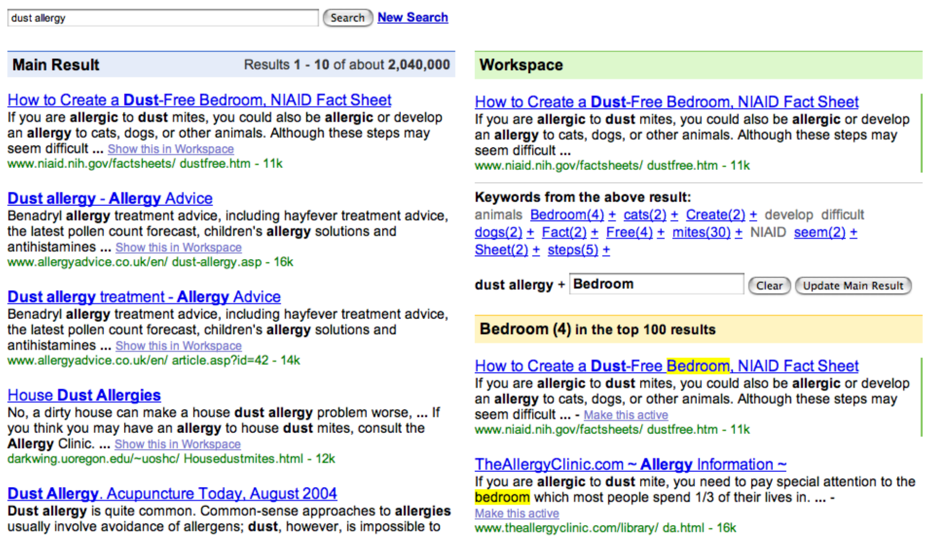

Fig. 1. Workspace interface

of the interface). Meanwhile, the document surrogates (title, snippets, URL, file size, etc.) of the 100 results were cached by using Javascript arrays. The rest of the processes was basically performed on the cached data until a new query was submitted.

The interface shows the *Workspace* area based on the pages searchers visited from the Main Result area. Alternatively, there was the *Show this in Workspace* link in the document surrogates as a shortcut to display a record in the workspace.

The workspace consisted of three parts (See Fig. 2). The top part was the click-through record. We called it an *active record*. The middle part listed a set of single words extracted from the active record's surrogates where stopwords (e.g., the, in, of, etc.) were excluded. The extracted words played the role of pseudo-facets in our interface. The number in the brackets indicated that how many other records shared the same keyword in the top 100 records. For example, there were four retrieved records whose surrogates contained the word *Bedroom*. The grey, non-hyperlinked words indicated that there was no other records which shared the same word (e.g., animals) in the 100 records. When one of the words was clicked, the word appeared in the query box of the middle part, and displayed a set of matched records in the bottom part of the workspace. The order of matched records were based on the original ranking in the Main Result. The pseudo-facet words were highlighted in the matched records to increase the awareness of contexts currently focused. A pseudo-facet word can be appended to the existing workspace query box by clicking the plus sign (+). When more than one words were added to the workspace query box, an intersection of the matched records were displayed in the bottom part (See Fig. 3(a)).

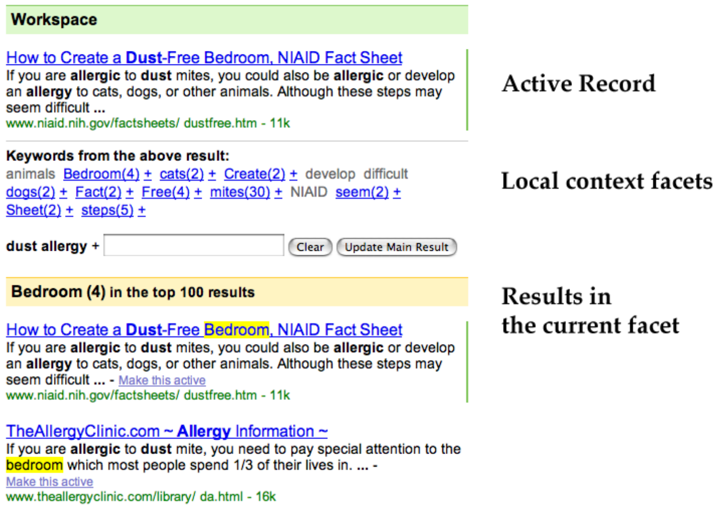

Fig. 2. Workspace Area (Query: dust allergy)

In addition, an incremental search function was implemented in the workspace query box. The incremental search allowed the interface to display a set of records matched any strings longer than two characters. The cached data enabled the interface to perform the incremental search on the 100 records in a fraction of second in most cases. Therefore, when the keywords from the active result were found to be not helpful or needed to modify, a searcher can change or input any words that might be useful for completing a search task (See Fig. 3(b)).

The workspace query box can also be used to submit a new query and get a new set of results in the Main Result area by clicking the *Update Main Result*

(a) Appending a new facet (mites) (b) Manually typing a new facet (relief)

Fig. 3. Further operations on facets

button. The existing query and words in the workspace query box would form a new query which would be then submitted to the search engine.

3 Experiment

A repeated measures within-participant design was used to compare the effectiveness of the proposed interface with the baseline interface. The independent variables were the system and task complexity. This section discusses the details of the experimental design adapted in our experiment.

3.1 Participants

A total of 24 people (2 females, 22 males) were recruited for our experiment. All participants but one were either the undergraduate students, postgraduate students, or research assistants of the University of Glasgow, UK. One participant was a visiting scholar of the University of Strathclyde, UK. The entry questionnaire established that the average age of participants was 25.4 (Min: 19; Max: 40;). Their major was in Computer Science (18), Electronic & Software Engineering (2), Chemical Engineering (1), Mechanical Engineering (1), Psychology (1), System Biology (1). Participants had on average 7.4 years of online search experience (Min: 4.0; Max: 10.0; SD: 1.6). Of 24, 21 performed several searches a day, two performed at least once a day, and one performed a couple of times a week. All indicated that they used Google most frequently.

3.2 Interfaces

Our proposed interface was compared to a baseline interface which only displayed the Main Result area. The workspace was disabled and the *Show this in Workspace* links were removed from the document surrogates in the Baseline interface. The same search engine with the same default search option (i.e., num = 100) was used in both interfaces to minimise the differences of response time and quality of search results. However, since we used the live Internet, no measure was taken to guarantee to receive an identical set of URLs for an identical query. In both interfaces, a cutoff for the maximum number of results was set to 100. As we will see in Section 4.2, this did not turn out to be a major restriction for participants. In the Main Result area, 10 results were shown per page, thus, participants could view up to 10 result pages per iteration. We used Firefox Web browser with a tab function in the experiment. When the title of pages was clicked, a new tab appeared to display the contents of the page.

3.3 Tasks

Participants were asked to carry out four search tasks in the experiment. As discussed in Section 1, one of our interests in this study was to evaluate the effectiveness of search interfaces with a different level of task complexity. A motivation for considering task complexity in the design and evaluation of search interfaces is that the level of task complexity can be controlled by an experimenter, to some extent, more systematically compared to other factors such as subject's background knowledge. One way to vary the task complexity is based on simulated work-task situations [15]. A simulated work task situation is "a semantically rather open description of the scenario and context of a given work task situation. Based on the scenario the test person formulates the search statement to the system (p.77)" [15].

Bell and Ruthven [4] proposed to control the task complexity by changing the amount of information on a task (e.g., information need, process, and outcome) provided in the simulated work-task situation form. They created a LC task by providing a greater amount of information in the situation form, and created a HC task by providing less. Using this approach, White, et al. [8] studied the effectiveness of implicit and explicit relevance feedback with three levels of task complexity. Their study shows that the effectiveness of relevance feedback is likely to be affected by task complexity.

Since it is not trivial to create a new set of tasks with a varied level of task complexity, it was decided to use the tasks which were shown to work. More specifically, four tasks were assembled from the work by Bell and Ruthven [4] and White, et al. [8]. While they originally designed three levels of task complexity for the same topic, in our experiment, we used only the highest complexity tasks and lowest complexity tasks to ensure the effect of complexity. The topics used in our experiment are 1) Dust allergy in workplace; 2) Music piracy on the Internet; 3) Petrol price; and finally, 4) Art galleries and museums in Rome.

3.4 Procedure

The user study was carried out in the following manner. At arrival time, participants were asked to read an information sheet which described the guideline for the participation and goal of the experiment. Upon the agreement of participation, participants were asked to fill in an entry questionnaire to indicate their age, gender, and search experience. Then they were presented with a training topic and explained the nature of simulated-work task. They were given approximately 10 minutes to familiarise with the search interfaces and task activity. This included a step-by-step tutorial for the two interfaces.

During the tasks, participants were asked to bookmark the web pages as an indication of finding perceived relevant pages. However, no explicit instruction was given to participants regarding the number of bookmarks required to complete the tasks. Participants were given up to 15 minutes to complete a task, but allowed to end it when they felt the task was completed. After the first task was completed, participants were asked to fill in a post-search questionnaire to provide subjective assessments about their search experience. Then a new task was given to them and change of interface was informed. The same procedure was repeated four times. The order of the systems and task complexity was rotated to minimise the bias of both effects. Participants were randomly assigned to either of the following rotations for task complexity (LC-HC-LC-HC or HC-LC-HC-LC) , and assigned to one of the following rotations for the systems (A-A-B-B, A-B-A-B, A-B-B-A, B-B-A-A, B-A-B-A, or B-A-A-B). This made a total of 12 possible combinations of the complexity levels and systems. Since it appeared to have little overlap among the search topics used in our study, their order was remained consistent across participants. This also allowed us to complete the experiment with a feasible number of participants.

After the completion of four tasks, participants were asked to fill in an exit questionnaire to indicate their overall preference of interfaces, followed by an open-end interview to capture their feedback and comments of the interfaces and experiment. User interactions were logged by the interface as well as a screen recorder software. An entire session took between 2 to 2.5 hours. Participants were rewarded with £15 for participation.

4 Results and analysis

This section presents the results of our study. We first investigate participants' perceptions regarding the search interfaces and tasks based on their subjective assessments. Then we examine the user interactions with the interface. Throughout the section, the Mann-Whitney U Test was used to establish the statistical significance of differences observed between the two interfaces and between the two levels of task complexity. The significance level was set to $p \leq .05$.

4.1 Participants' perceptions

As an overall assessment of the interfaces, participants were asked to indicate the preference of two interfaces based on their experience of four search tasks at the end of experiment. 21 out of 24 (87.5%) indicated that they preferred the Workspace interface over the Baseline interface. This suggests that most participants appeared to welcome the functionality offered by the workspace. This section investigates participants' perceptions of the interfaces and search tasks based on their subjective assessments established by the post-search questionnaires. A 7-point scale was used to capture participants' assessments where a lower score represented a more positive assessment in the analysis. In the following tables, the figures are the mean value of 24 sessions, and the numbers in the brackets represent the standard deviation (SD), unless otherwise stated.

Table 1 shows participants' assessments on their satisfaction with the outcome of search tasks. Overall, participants appeared to feel more satisfied with

Table 1. Search task satisfaction (Scale 1-7; Lower = Better)

	Baseline	**Workspace**	**All**
Low Complexity	3.08 (1.64)	2.42 (1.02)	2.75 (1.39)
High Complexity	2.71 (1.65)	2.38 (0.97)	2.54 (1.35)
All	2.90 (1.64)	2.40 (0.98)	2.65 (1.37)

the high complexity (HC) tasks than lower complexity (LC) tasks. For both levels of task complexity, participants appeared to feel more satisfied with the outcome of search tasks when the Workspace interface was used. While the difference is not statistically significant, the pattern was consistent across the four tasks for both levels of complexity except Task 4 (LC & HC).

Table 2 shows participants' assessment on the usefulness of the interfaces. As can be seen, participants indicated a more positive assessment on the interfaces in the HC tasks than LC tasks. The assessments on the Workspace were consistently more positive than the Baseline in both levels of complexity except Task 1 (HC). The difference between the two interfaces was statistically significant in both levels of complexity. This suggests that participants found the Workspace interface more useful than the Baseline interface.

The tests discussed so far seem to suggest that participants generally had a positive assessment on the Workspace interface. One of the features in the Workspace interface was that participants were given a control on how the search results were organised and browsed. The next two tests investigate how this feature affects participants' assessments on finding relevant pages.

Table 3 shows participants' assessment on how easy it was to identify relevant pages in the search results. In LC tasks, participants appeared to find the Baseline interface easier to identify relevant pages while in HC tasks the Workspace interface was given an overall positive assessment. The pattern was less consistent across the tasks compared to the previous tests. Participants indicated that they found it easier to identify relevant pages in the Baseline interface in Task 3 (HC) and Task 4 (LC & HC). No difference was statistically significant.

Table 4 shows participants' assessment on how easy it was to find new information that participants had not already seen before. Overall, participants appeared to find it harder to search for new information than relevant pages. While the assessment on the Baseline interface was less positive in HC tasks than LC tasks, the relation was in inverse in the Workspace interface. In Task 3 and 4, the assessments of the Workspace interface were consistently more positive than the Baseline interface, while there were ties in the rest of tasks across the complexity. However, no difference was statistically different.

Table 2. Usefulness of Interface (Scale 1-7; Lower = Better)

	Baseline	**Workspace**	**All**
Low Complexity	3.67 (1.47)	2.46 (1.14)	3.06 (1.44)
High Complexity	3.08 (1.41)	2.21 (1.10)	2.65 (1.33)
All	3.38 (1.45)	2.33 (1.12)	2.85 (1.39)

Table 3. Identifying relevant pages (Scale 1-7; Lower = Better)

	Baseline	**Workspace**	**All**
Low Complexity	3.00 (1.45)	3.08 (1.28)	3.04 (1.35)
High Complexity	2.92 (1.67)	2.50 (1.22)	2.71 (1.46)
All	2.96 (1.54)	2.79 (1.27)	2.88 (1.41)

Finally, we asked participants to indicate their motivations for accessing the workspace area during the search tasks. Table 5 shows the result of three motivations. As can be seen, browsing of the search results appeared to be the overall strongest motivation for accessing the workspace area while focusing on a set of pages based on particular keyword(s) also appeared to be a popular reason. It is interesting to see that the strongest motivation is different across the complexity.

Summary: This section investigated participants' perceptions on the search tasks and interfaces from several aspects. Overall, participants' assessments on the Workspace interface tended to be positive, but the results showed some degree of variance across the search tasks. Perhaps, the most consistent result is that the SD of the Workspace interface was lower than the Baseline interface in most tests. Therefore, it is likely that a positive assessment can be more consistent across participants in the Workspace interface than the Baseline interface.

4.2 User interactions

The last test indicated that participants might access to the workspace area for a different motivation across the task complexity. This section presents the user interactions with the interfaces.

Table 6 shows the user interactions with the workspace area. The second column represents the number of the shortcut actions taken in the sessions. The shortcut was a way to display a result in the workspace area without visiting the URL. The third column represents the number of pseudo-facet selections made to display a particular groups of search results. The fourth column represents how often a new facet was appended to the current set. Finally, the fifth column represents the manual editing of pseudo-facets by participants.

Table 4. Finding new information (Scale 1-7; Lower = Better)

	Baseline	**Workspace**	**All**
Low Complexity	4.08 (1.59)	3.71 (1.33)	3.90 (1.46)
High Complexity	4.25 (1.78)	3.63 (1.66)	3.94 (1.73)
All	4.17 (1.67)	3.67 (1.49)	3.92 (1.59)

Table 5. Motivations of accessing workspace (Scale 1-7; Lower = Stronger)

	Browse	**Focus**	**Weed-out**
Low Complexity	2.71 (1.85)	2.67 (1.62)	4.00 (2.00)
High Complexity	2.14 (0.94)	2.82 (1.50)	3.73 (2.07)
All	2.42 (1.47)	2.74 (1.54)	3.86 (2.02)

Table 6. Access to the workplace area

	Shortcut	**Facet select**	**Facet add**	**Facet edit**
Low Complexity	4.13 (4.41)	3.83 (4.41)	2.42 (3.46)	2.04 (2.93)
High Complexity	4.29 (5.21)	4.54 (4.12)	1.67 (2.43)	1.29 (1.27)
All	4.21 (4.78)	4.19 (4.24)	2.04 (2.98)	1.67 (2.26)

As can be seen, participants appeared to modify existing pseudo-facets more actively in LC tasks than HC tasks. This indicates that participants were more likely to have particular keywords in mind during the LC tasks. On the other hand, participants seemed to be selecting extracted pseudo-facets to browse the search results in HC tasks. This appears to echo the result shown in Table 5. The following analysis investigates how these accesses to the workspace affect other aspects of information seeking process.

Table 7 shows the number of iterations carried out during the search sessions. As can be seen, participants appeared to submit fewer queries in HC tasks compared to LC tasks in the Workspace interface while the number was similar in the Baseline interface. The difference between the two interfaces was statistically significant in HC tasks. This suggests that accessing the workspace can lead to lowering the amount of effort otherwise required to reformulate existing queries. Note that the number of iterations performed by participants is larger than the average search engine users. This can be due to the fact that participants were engaging in the *informational search* [1] as opposed to the navigational search frequently found in the log studies such as [16].

Table 8 shows the number of pages participants clicked to visit during the search sessions. While the frequency of click-through in the Baseline interface appeared to be increased in HC tasks, the difference between the complexity was remained to be small in the Workspace interface. The overall difference between the two interfaces was small.

Fig. 4 shows the distribution of the original rank positions of click-through pages. As can be seen, participants' click-through is concentrated on the records in the first few pages with the Baseline interface. On the other hand, the records from a wider range of ranks were visited by participants when the Workspace interface was used. It also seemed to reduce the click-through from the first page of search results. The following tests addresses how this difference affects participants' relevance assessments.

During the search tasks, participants were asked to bookmark the web pages which were perceived to be relevant. Table 9 shows the number of pages bookmarked. As can be seen, overall, participants saved a larger number of pages

Table 7. Number of iterations

	Baseline	**Workspace**	**All**
Low Complexity	8.46 (6.45)	8.33 (5.17)	8.40 (5.78)
High Complexity	8.42 (5.56)	6.00 (4.87)	7.21 (5.32)
All	8.44 (5.96)	7.17 (5.11)	7.80 (5.56)

Table 8. Number of click-through URLs

	Baseline	**Workspace**	**All**
Low Complexity	11.79 (7.89)	12.13 (5.89)	11.96 (6.89)
High Complexity	13.25 (7.05)	12.00 (6.59)	12.63 (6.78)
All	12.52 (7.44)	12.06 (6.19)	12.29 (6.81)

in HC tasks compared to LC tasks. The difference between the two levels of complexity was statistically significant in both systems. This suggests that participants felt that they needed to find more relevant pages to complete HC tasks than LC tasks. This is not surprising given that participants had a higher level of uncertainty in HC tasks. The task-breakdown of the results suggest that the largest difference between the two interfaces was in Task 1 (HC) where on average 11 pages were bookmarked with the Baseline interface while 4 pages were saved with the Workspace interface. However, the difference was not statistically significant.

Table 10 shows the successful click-through rate (SCTR) which is the percentage of click-through pages that lead participants to bookmark a page (i.e., find a perceived relevant page). As can be seen, the Baseline interface appeared to offer a higher SCTR than the Workspace interface across the complexity. The task breakdown of the result suggests that participants tended to achieve a higher SCTR in three topics in HC task, and two topics in LC tasks. While the

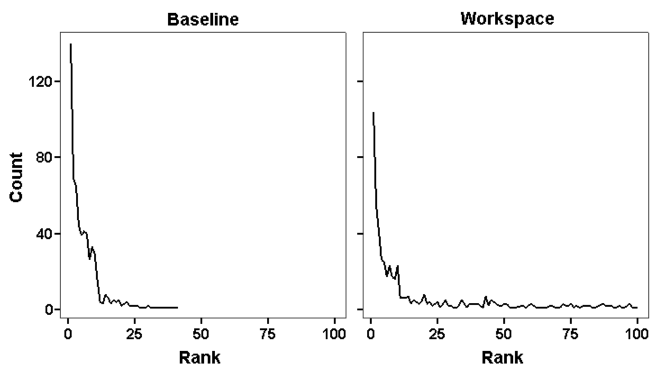

Fig. 4. Distribution of click-through pages ranking

Table 9. Number of bookmarked pages

	Baseline	Workspace	All
Low Complexity	5.79 (5.45)	6.08 (4.46)	5.94 (4.93)
High Complexity	8.50 (6.07)	6.88 (4.38)	7.69 (5.29)
All	7.15 (5.86)	6.48 (4.38)	6.81 (5.16)

Table 10. The successful click-through rate (%)

	Baseline	Workspace	All
Low Complexity	53.3 (28.3)	50.3 (23.3)	51.8 (25.7)
High Complexity	63.3 (26.0)	59.3 (19.7)	61.3 (22.9)
All	58.3 (27.3)	54.8 (21.8)	56.6 (24.7)

difference was not statistically significant, this suggests that participants might find perceived relevant pages more accurately with the Baseline interface.

Fig. 5 shows the box plot based on the same data used in Table 10 (Note that the middle bar in the box is the Median). As indicated by the SD of Table 10, in HC tasks, a wider range of variance was found in the Baseline interface compared to the Workspace interface. For example there is one Baseline session with no page bookmarked in each level of task complexity. Therefore, one of the characteristics of the Workplace Interface seems to be the stability of performance among participants. However, the range of SCTR appears to be similar in LC tasks.

Finally, Table 11 shows the time taken to complete the search tasks. Participants appeared to take longer to complete the tasks when the Workspace interface was used. The pattern was consistent across the complexity but the difference was not statistically significant.

Summary: This section presented the results regarding the user interactions with the search interfaces. The result suggests that a respectable number of actions was taken place in the workspace area. It also indicates that the use of the workspace area is likely to vary over the task complexity. The use of the

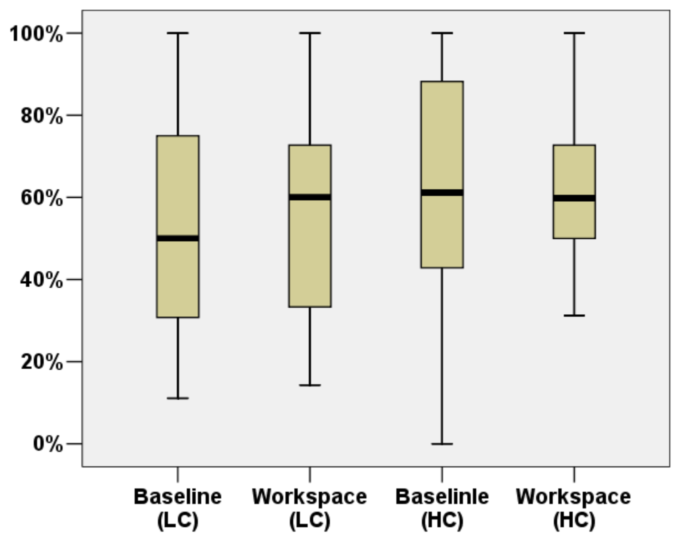

Fig. 5. The successful click-through rate (%)

Table 11. Task completion Time (min)

	Baseline	**Workspace**	**All**
Low Complexity	10.82 (5.00)	12.56 (3.50)	11.69 (4.36)
High Complexity	11.33 (4.15)	11.47 (3.89)	11.40 (3.98)
All	11.07 (4.55)	12.02 (3.70)	11.55 (4.16)

workspace areas is likely to reduce the number of iterations in HC tasks. However, participants might find more perceived relevant pages with the Baseline interface. The SCTR also appeared to be higher with the Baseline interface, but the result was inconclusive.

5 Implications

Our experimental results have several implications for the design of search interfaces, and the evaluation methodology for IIR systems using HC tasks.

5.1 Search interface design

The faceted grouping of search results is an interesting alternative to clustering techniques. However, the implementation of existing faceted applications appears to reply on the sufficient knowledge of collection structure, thus, they can be expensive to develop and maintain for a new collection. In this study, we proposed a search interface which allowed searchers to slice and dice the information space using local contexts. One of the implications in this study is that the local contexts can be exploited for the faceted-like interactions to explore the search results. Our approach did not require an extensive analysis of the collection to offer a flexible way of exploring search results. The feedback from participants was more encouraging than we had expected. One of the reasons for participants' positive assessments on the proposed interface appears to be due to the workspace area. The interface offered the area so that searchers can explore search results without losing the initial set. Therefore, our work supports the proposal and findings described in [9] and [14] for the importance of independent workspace area in IIR. A limitation of the current implementation of our interface is that a range of pseudo-facets extracted from the local contexts is sometimes limited. We are interested in the use of top ranking sentences (TRS) [17] to address this issue in our future work. The TRS can be used to increase the awareness of active records as well as the range of pseudo-facets offered to searchers. We are particularly interested in investigating the effect of TRS in the successful click-through rate.

5.2 Evaluation methodology

Our user study was based on the simulated work task situation with varied level of task complexity. Overall, this framework provided us with a good experimental design to evaluate the effectiveness of our approach. On the other hand, the experience of the framework gave us several ideas of the further development, especially in the context of high complexity tasks. A high level of task complexity means a higher degree of uncertainty involved in search tasks. However, we noticed that sometimes the uncertainty becomes more freedom in participants' perceptions of tasks, thus, a HC task can be perceived as an easier task. The higher successful click-through rate shown in Table 11 appears to support our speculation. This indicates at least two things. First of all, we need a further development on the protocol for using the simulated work task situation with HC tasks. Another is the potential benefit of setting a goal as a part of the task situation. The current framework is designed to control the level of complexity based on the task inputs, processes, and outputs [4]. Setting varied levels of search goals might be used complementary in the framework. An example of task goals might be "You can end your search when you can confidently tell the experimenter if the increase of X is related to the problems of Y. You must show appropriate evidences that support your conclusion." An appropriate evidence might have the properties such as credibility, timeliness, and/or locality. This type of goals appears to reflect the work tasks in organisations, thus, an interesting venue to apply the proposed search interface.

6 Conclusion

This paper proposed and evaluated an approach to Web search and browsing based on local contexts of query terms. A task-based study with varied levels of task complexity suggests that the local contexts can be exploited as the source of search results browsing in IIR systems. The overall positive assessments on the proposed interface suggests that there are avenue for further development and research on our approach. We also discussed potential ideas to develop the evaluation methodology using high complexity tasks based on our experience in this study.

One of the limitations would be the range of topics used in this study. While our evaluation was based on four different search tasks, the number of topics tested in the experiment was relatively limited. A further study with additional domains will give us a comprehensive picture of the system performance presented in this paper. Also, we used a single search engine, and no measure was taken to ensure that an identical set of URLs was retrieved in response to the same query during the experiment.

7 Acknowledgements

The authors thank the participants of our experiment for their time and invaluable feedback. Our thanks also go to Prof. Stephen A. Brewster for his helpful feedback on the design of our interface. The authors also thank the anonymous reviewers for their helpful suggestions on the paper. This work was funded by the EPSRC (Grant ref: EP/C004108/1). Any opinions, findings, and conclusions described here are the authors and do not necessarily reflect those of the sponsor.

References

1. Broder, A.: A taxonomy of web search. SIGIR Forum **36**(2) (2002) 3–10

2. Rose, D.E., Levinson, D.: Understanding user goals in web search. In: Proceedings of the 13th international conference on World Wide Web, New York, NY, ACM (2004) 13–19
3. Byström, K., Järvelin, K.: Task complexity affects information seeking and use. Information Processing & Management **31**(2) (1995) 191–213
4. Bell, D.J., Ruthven, I.: Searchers' assessments of task complexity for web searching. In McDonald, S., Tait, J., eds.: Advances in Information Retrieval, 26th European Conference on Information Retrieval. Lecture Notes in Computer Science 2997, Sunderland, UK, Springer (2004) 57–71
5. Marchionini, G.: Information Seeking in Electronic Environments. Cambridge University Press, Cambridge, UK (1995)
6. Shiri, A.A., Revie, C.: The effects of topic complexity and familiarity on cognitive and physical moves in a thesaurus-enhanced search environment. Journal of Information Science **29**(6) (2003) 517–526
7. Vakkari, P.: Task complexity, problem structure and information actions integrating studies on information seeking and retrieval. Information Processing & Management **35**(6) (1995) 819–837
8. White, R.W., Ruthven, I., Jose, J.M.: A study of factors affecting the utility of implicit relevance feedback. In: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil, ACM (2005) 35–42
9. White, R.W., Kules, B., Drucker, S., schraefel, m.c.: Supporting exploratory search: A special issue of the communications of the acm. Communications of the ACM **49**(4) (2006) 36–39
10. Hearst, M.A., Pederson, J.O.: Re-examining the cluster hypothesis: Scatter/gather on retrieval results. In Frei, H.P., Harman, D., Schable, P., Wilkinson, R., eds.: Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Zurich, Switzerland, ACM (1996) 76–84
11. Zamir, O., Etzioni, O.: Grouper: A dynamic clustering interface to web search results. Computer Networks **31** (1999) 11–16
12. Zhang, J., Marchionini, G.: Evaluation and evolution of a browse and search interface: relation browser. In: Proceedings of the 2005 national conference on Digital government research. (2005) 179–188
13. Yee, K.P., Swearingen, K., Li, K., Hearst, M.: Faceted metadata for image search and browsing. In: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM (2003) 401–408
14. Urban, J., Jose, J.M.: Can a workspace help to overcome the query formulation problem in image retrieval? In: Advances in Information Retrieval, 28th European Conference on Information Retrieval, London, UK, Springer (2006) 385–396
15. Borlund, P.: Experimental components for the evaluation of interactive information retrieval systems. Journal of Documentation **56**(1) (2000) 71–90
16. Jansen, B.J., Spink, A., Bateman, J., Saracevic, T.: Real life information retrieval: A study of user queries on the web. ACM SIGIR Forum: A Publication of the Special Interest Group on Information Retrieval **32**(1) (1998) 5–17
17. Tombros, A., Sanderson, M.: Advantages of query-biased summaries in information retrieval. In Croft, B.W., Moffat, A., van Rijsbergen, C.J., eds.: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Melbourne, Australia, ACM (1998) 2–10