Effectiveness of Additional Representations for the Search Result Presentation on the Web[*]

Hideo Joho and Joemon M. Jose

Department of Computing Science

University of Glasgow

`hideo,jj@dcs.gla.ac.uk`

Abstract

The presentation of search results on the web has been dominated by the textual form of document representation. On the other hand, the document's visual aspects such as the layout, colour scheme, or presence of images have been studied in a limited context with regard to their effectiveness of search result presentation. This article presents a comparative evaluation of textual and visual forms of document representation as *additional* components of document surrogates. A total of twenty-four people were recruited for our task-based user study. The experimental results suggest that an increased level of document representation available in the search results can facilitate users' interaction with a search interface. The results also suggest that the two forms of additional representations are likely beneficial to users' information searching process in different contexts.

1 Introduction

Representation is one of the fundamental issues in the research and development of information retrieval (IR) systems. The IR processes such as indexing, ranking, and query matching usually rely on the representation which is derived from the original objects (e.g., documents, images, videos). For example, a vector space model represents a document and query based on multi-dimensional vectors (Salton, 1971). A probabilistic model represents them based on a probability of occurrence of terms (Robertson et al., 1995). Low level features such as colour histograms are also used in image retrieval (e.g., Gupta and Jain, 1997). These alternative representations of the original objects enable an IR system to offer retrieval on a large scale collection of information objects. The World Wide Web (or the web) is one such environment where IR systems have been effectively used to find relevant information.

Representation of the original objects is also an important issue for the presentation of retrieved results. Just like an inappropriate representation of information might cause an IR system to retrieve irrelevant objects, an inadequate representation of retrieved objects might cause misleading relevance assessments on search result, thus, it might degrade the potential effectiveness of underlying retrieval models. In IR, an alternative representation of the retrieved objects is sometimes referred to as a *surrogate*. For example, the document surrogate used in search engines typically consists of a title, snippet (relevant text fragment), URL, size, and/or file type of the original web page. Metadata such as a category description or classification code of retrieved documents have also been used as a part of surrogates (Shiri and Revie, 2003). A version of the Okapi system offers query terms' frequency of occurrence in retrieved documents as supplementary data (Beaulieu and Gatford, 1998).

However, the study on the effectiveness of surrogates is relatively limited compared to the extensive research that has been carried out on the retrieval models. Much research has investigated the approaches to organising, summarising, or visualising retrieved objects (e.g., Chen and Dumais, 2000; Hearst and Pederson, 1996; Zamir and Etzioni, 1999). While these techniques aim to help users find relevant documents from search results, they do not necessarily address questions such as "What constitutes a good surrogate?", "To what extent does the effectiveness of surrogates depend on a task at hand?", "Is there room for improving the surrogates currently used by search engines?". The study presented in this article aims to address some of these questions based on four different search interfaces designed for our experiment.

This work is motivated by two aspects of document surrogates which are commonly found in search engines. The first aspect is the poly-representation of retrieved objects.

In the development of cognitive IR models, Ingwersen (1992) discussed the importance of representing an object using multiple forms in all levels of user interactions with IR systems. Recent empirical studies have suggested that the poly-representation can be integrated into a retrieval model (Larson, 2005), or can be used as a means of facilitating query formulation processes (Kelly et al., 2005). Evidently, document surrogates used by existing IR systems are one of the examples of poly-representation. Searchers are given a control on what aspect(s) of retrieved objects to use for their relevance assessments on search results. This also suggests that one way to improve the presentation of search results might be to augment existing surrogates with a different representation derived from the retrieved objects.

Another aspect is the lack of visual elements exploited as a component of document surrogates. The texts and document structures have been extensively exploited for effective retrieval models and search result presentations. This trend is to some extent still evident on the web. However, web pages contain a wider range of attributes than the conventional text documents, including multiple colours, layouts, and images. These visual elements of documents are likely to have an effect in the search process (Tombros et al., 2005), thus, they can be exploited as a component of surrogates.

The main contribution of this article is two-fold. Firstly, the effectiveness of textual and visual forms of document representation is evaluated as additional components of document surrogates. The four search interfaces devised in this work enable us to evaluate the two forms (i.e., textual and visual) of representations in a systematic way. Secondly, both quantitative and qualitative analysis are carried out to evaluate the effectiveness of the proposed search interfaces based on a task-based user study. In our experiment, participants search for relevant information on the live Internet using their own queries. With a realistic laboratory condition, it is anticipated that our study provides further insight into the effectiveness of search result presentations and their impact on other aspects of user's information seeking behaviour on the web.

The rest of the article is organised as follows. The next section reviews the existing work on augmenting surrogates with textual and visual representations. Section 3 presents the details of the four search interfaces designed for our experiment. Section 4 discusses the evaluation methodology adopted in our study. Section 5 presents the experimental results and analysis. Section 6 discusses the implications of our findings on the interface design of IR systems. Finally, Section 7 concludes the article and future work is discussed.

2 Augmentation of document surrogates

Many examples of search result presentations of IR systems can be found in the literature. For example, Beaulieu (1997) discussed the development of user interfaces of the Okapi system from the period where a character-based presentation was built upon a library catalogue and other bibliographic databases (early 90s) to a more recent version which offered a full-text search on a larger full-text collection. We can see from the examples shown in Beaulieu's paper that the document surrogates tend to consist of a title, author, and publication year extracted from the collection. Another version of the Okapi system which is built upon a TREC collection (newspapers) presents a document ID, publication date, title (if any), and the first few lines of the document (Beaulieu and Gatford, 1998). It should be noted that the surrogates used in search engines are not radically different from these, suggesting that there might be a consensus regarding what should constitute a surrogate among the researchers and developers of IR systems.

Greene et al. (2000) discuss the roles of surrogates in search result presentations from two perspectives: previews and overviews. They argue that "an effective preview is an information surrogate that communicates to the user, at the appropriate time, sufficient information about the primary object it represents to support users in making a correct judgement about the relevance of that object to the user's information need" (p. 381). The overview, on the other hand, is used to represent collections of objects. Their paper discussed the design of previews and overviews in search results based not only on a text-based application but also other media such as images and movies in the context of digital libraries. A key issue behind the construction of effective surrogates appears to be the poly-representation of retrieved objects (Ingwersen, 1992). Multiple forms of objects' representation allow an IR system to present search results in a wider range of contexts. This is particularly important for an IR system to be adaptive to the searchers who are likely to have varied knowledge and needs in the task at hand. The poly-representation also gives the searchers a greater control on how they assess the relevance of retrieved objects presented in search results.

One of the techniques which has been shown to help document's surrogates being adaptive is a query-biased summarisation (Tombros and Sanderson, 1998). Unlike the first few lines of documents which are static to user's information needs, a query-biased summary offers a greater level of contexts regarding how query terms are used in the documents. Empirical studies have suggested that the query-biased sentences (called top ranking sentences or *TRS*) can facilitate relevance assessments in search results (Tombros and Sanderson, 1998), and that the sentences can be a more effective representation than fragment of texts often used as search engines' snippets (White et al.,

2003b). However, the effectiveness of TRS has been evaluated as a replacement of existing document's surrogate, thus, there is a scope for further studies in which they are evaluated as an additional representation of retrieved objects. In addition, the TRS appears to be an interesting candidate for the textual form of representation in our study since they are likely to offer an additional context of query terms which might not be provided by the existing surrogates of search engines.

A distinct approach to augment existing surrogates is based on the visual elements of retrieved objects. While the textual representations still dominate the search result presentations on the web, it has been suggested that the searchers also use the visual elements in the relevance assessments of web pages (Tombros et al., 2005). The colour schemes, page layouts and presence of images are likely to offer a different type of contexts which are not necessarily conveyed by textual representations. In particular, there has been a growing interest in exploiting the thumbnail images of web pages to support user's information searching process (e.g. Dziadosz and Chandrasekar, 2002; Lam and Baudiscch, 2005; Woodruff et al., 2002). A good overview of various designs and usages of thumbnail images in IR applications is provided by Woodruff et al. (2002). In their paper, the authors argue that thumbnails have been studied in a limited scope such as searcher's memory aid of previously visited pages.

The studies by Dziadosz and Chandrasekar (2002) and Woodruff et al. (2002) have extended such a scope to evaluate the effectiveness of thumbnails as document surrogates in search result presentations. For example, Dziadosz and Chandrasekar (2002) examined if the presence of thumbnails in surrogates facilitated searcher's relevance assessments of unseen retrieved web pages. Their experiment suggests that thumbnails are likely to speed up relevance assessments, but it could also increase the rate of false positive judgements. In a mobile-device environment, a thumbnail view of web pages has been shown to be more appropriate for the smaller screen (Lam and Baudiscch, 2005). Woodruff et al. (2002) developed an enhanced thumbnail which allows searchers to view magnified relevant texts in a thumbnail.

While some of these applications currently require additional software to use, it has suggested the potential usefulness of thumbnails in supporting user's information searching process. However, since thumbnails are a relatively new media to convey information about web pages compared to text-based applications, we argue that there is still room for investigating their effectiveness. For example, participants of the existing studies were often asked to engage only in browsing and judging of retrieved pages. Therefore, there is room for investigating the effectiveness of thumbnails in a more holistic condition where the searchers engage in all aspects of searches. Also, the

effectiveness of thumbnails was typically measured by task completion time and number of pages visited from search results. This suggests that the qualitative analysis of experimental results is relatively limited.

This article presents a comparative evaluation of the textual (TRS) and visual (thumbnail images) forms of document representation as additional components of document surrogates. Participants of our experiment engage in all aspects of search tasks. Both quantitative and qualitative aspects of experimental results are equally emphasised in our analysis. We aim to achieve these objectives by designing four search interfaces which are discussed in the next section.

3 Interfaces for search result presentation

We augment the web search engine Google with three new result layouts. Our interfaces collected user queries and forward to Google using their API[1]. The result list from Google was then collected and processed. The information needed for the new interfaces are created at this time. Like in the web search result pages, the users could peruse ten document records at a time. After this, they could either reformulate the query or peruse the next ten records. The rest of this section presents the four layouts used in our experiment.

Layout 1: Baseline   The baseline layout was designed to provide an almost identical interface to the search result of Google. For each record, it presented a title, snippet, URL, size, and the hyperlinks of cached page and similar page.

Layout 2: Baseline + TRS   The second layout integrated the top ranking sentences (TRS) into the baseline layout. We used a version of TRS generation software developed and evaluated by White (2004). The software ranked sentences based on a mixture of factors such as term weight, sentence location in a document, and HTML tags. The detail of the implementation is found in White (2004). The top ranked sentences were inserted below the snippet as a list, and background was highlighted to clarify the distinction between the snippet and TRS. Up to three sentences were shown per result record. The query terms were highlighted in bold in the same manner as it would have been in the title and snippet. There was some run-time overhead in generating TRS for retrieved documents. To minimise the difference of response time among the layouts, TRS was always created when a new query was submitted to the interface, but only displayed in

---

[1] Available from `www.google.com/apis`.

Layout 2 and 4.

Layout 3: Baseline + Thumbnail   The third layout integrated a thumbnail image of the web page's screenshot into the baseline layout. The thumbnails were fetched from the Alexa's thumbnail archive[2]. The thumbnail was placed on the left side of the other document surrogates, and it was linked to the URL of the page. The size of thumbnails was 112 (width) and 82 (height) pixels which was perhaps too small to read the texts, but we considered that it was large enough to grasp the visual aspects of pages such as the layout, colour scheme, or the presence of images.

Layout 4: Baseline + TRS + Thumbnail   The last layout was the combination of Layout 2 and 3 (See Figure 1). While this layout took up the largest space in the screen, it was designed to provide the largest amount of information per record among the four layouts.

Figure 1: Search result with TRS and thumbnail (Layout 4).

As can be seen, we designed the four layouts where the different variable between them remained to be a single element. We considered that it was important to minimise the difference between layouts to evaluate the effectiveness of TRS and thumbnails in a meaningful way. The next section presents the details of our experiment based on the four layouts.

4   Experiment

A comparative user study was carried out to evaluate the effectiveness of the four search result layouts described above. This section discusses the evaluation methodology adopted in our experiment.

---

[2]The Alexa archive (`www.alexa.com`) did not always contain the thumbnail of the web pages retrieved during our experiment. Our understanding was that a missing thumbnail was replaced by a parent site when it was available. Otherwise it showed the Alexa's logo image to indicate the absence of thumbnails. In our experiment, a missing thumbnail was treated as a similar case to a dead link on the real-life web.

4.1 Participants

A total of twenty-four people (6 female and 18 male) were recruited for our experiment. 23 of the participants were postgraduate students of the University of Glasgow, and the remaining one was an affiliated member of the University. The recruitment was carried out through our call for participation distributed through University mailing lists and subsequent word-of-mouth fashion. The first 24 people who contacted us were recruited. The entry questionnaire established that the range of age varied from 20 to 37 with an average of 27.7. Their experience with search engines varied from 4.5 to 11 years with an average of 7.1 years. All participants carried out several searches every day, and 22 (91.6%) of them used Google most frequently.

4.2 Tasks

Participants were asked to carry out four search tasks in the experiment. One of our research interests was to evaluate the effectiveness of the proposed interfaces based on a range of search tasks. The tasks were designed based on the simulated work task situation framework (Borlund, 2000). The framework described a task as a form of short scenario. The scenario explained the contexts and motivation of the search with sufficient information about the relevance of pages. An overview of the tasks used in our experiment is as follows.

Task 1: Background search task This task asked participants to find general background information on a topic. In our experiment, participants were asked to find the pages which provide information about the recent change of student populations.

Task 2: Decision-making task This task asked participants to make a decision about a topic. In our experiment, participants were asked to find the best Hi-Fi speakers available in a target price. Participants were encouraged to compare the speakers' details in the decision making process. Task 1 and 2 were based on the descriptions originally proposed by Tombros et al. (2005).

Task 3: Known item task This task asked participants to find information about a topic which was previously known. In our experiment, participants were asked to find the current whereabouts of a person who was assumed to be a previous colleague of the searcher. However, all participants were asked to find the same person.

Task 4: Topic distillation task This task asked participants to find a list of key resources for a topic. The definition of key resources was based on the instruction of the Web Track of TREC[3]. The main criteria for being a key resource was that the website was principally devoted to the topic. In our experiment, participants were asked to find the key resources for designer handbags.

4.3 Procedures

The user study was carried out in the following manner. At arrival time participants were asked to read an information sheet which described an overview of the experiment and guideline for the participation. Upon the agreement of participation, participants were asked to fill in an entry questionnaire to indicate their background information. 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. During the training session, the four layouts were introduced to participants and questions regarding the interface and tasks were answered.

During the tasks, participants were asked to bookmark the pages which they thought relevant to the tasks. However, no explicit instruction was given to participants regarding the number of bookmarks required to complete the tasks. We asked participants to bookmark pages to ensure their engagement in the search tasks. All participants have used the bookmark of web browsers in the past and they did not express any difficulty of bookmarking during the experiment. Participants were given up to 15 minutes to complete a task, but were allowed to end it when they felt they had completed the tasks.

After the first task was completed, participants were asked to fill in a post-search questionnaire to provide subjective assessments about their search. A new task was then given to them and they were informed about the change of layout. The same procedure was repeated four times. Each participant carried out all four tasks using a different order of the four layouts. The Greaco-Latin square design (Tague-Sutcliffe, 1992) was used to control participants' learning effect between tasks and layouts. When the four tasks were completed, participants were asked to fill in an exit questionnaire to indicate their overall preference of layouts, followed by an open-ended interview to capture their feedback and comments about the result presentation and experiment. The whole session tended to take between 1.5 to 2.5 hours. Participants were rewarded with $25 for their participation.

---

[3]Available from es.csiro.au/TRECWeb/guidelines_2004.html.

5   Results

This section presents the experimental results of our study based on the 96 search sessions carried out by 24 participants. Participants were divided into two groups based on their search experience established by the entry questionnaire. The less experienced group (denoted as *Less Exp.* in the tables) had on average 5.5 years of search experience, while the more experienced group (denoted as *More Exp.*) had on average 8.7 years of search experience. Overall, the more experienced group had the majority of participants from the Computer Science field.

We ran the Friedman Test (non-parametric, dependent, n>2) to establish the statistical significance of the differences found among the four layouts. We also ran the Mann-Whitney U Test (non-parametric, independent, one-tailed, n=2) to establish the significance of the difference between the two experience groups. The Mann-Whitney U Test was also used for the difference between Layout 1 and the cumulate of Layout 2, 3, and 4 data. The motivation behind the latter test was to investigate an overall effect of the layouts where *something* (i.e., TRS, thumbnail, or both) was added, that is, the level of document representation was increased in the presentation of search results. The difference between Layout 1 and the cumulated of Layout 2-4 was tested based on all participants, unless otherwise stated. The significance level was set to $p \leq .05$ for all tests.

This section consists of three subsections. Firstly, the result of user interaction with the four layouts is presented. The data presented in this section was recorded by our system during the experiment. Secondly, the results of participants' perceptions on the interface and search experience are presented. The data presented in this section was established by the post-search questionnaires. Lastly, the result of participants' system preference and feedback on the layouts is presented. The data presented in this section was based on the mixture of the exit-questionnaire and interview carried out at the end of the experiment.

5.1   User interactions

The result of user interactions is presented by three aspects as follows: query formulation, browsing and click-through, and task completion time. In the tables, the mean value of the samples is presented along with their standard deviation given in the brackets. The size of samples used for the analysis is given at the bottom of the tables.

5.1.1 Query formulation

Table 1 shows the number of queries submitted to the interfaces during the tasks and their details such as the number of unique words and query length. As can be seen, participants appeared to submit more queries in Layout 2 to 4 compared to Layout 1. While no statistical significance was found by the Friedman Test, the difference between Layout 1 and the cumulate of Layout 2-4 was found to be significant by the Mann-Whitney U Test ($Z = -1.728, p = .042$). Therefore, there seems to be a relationship between the level of document representation and the number of queries submitted to the interface. There can be two possible interpretations for this. A positive interpretation is that the additional representation of documents facilitated participants' query reformulation, since they were given an increased amount of information about retrieved documents. A negative interpretation is that additional representations were misleading for participants, thus, they had to submit more queries to complete a task. If the latter was the case, it was likely that participants with Layout 2 to 4 took a longer time to complete the tasks than Layout 1. As we see in Section 5.1.3, however, we did not find a significant difference in the task completion time across the layouts. Therefore, *it is likely that an increased level of document representation shown in the search results can facilitate users' query formulation process*. We did not find the statistical significance of the difference between the experience groups. 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* (Broder, 2002) as opposed to the navigational search frequently found in the log studies such as Jansen et al. (1998)

The range of words used during the tasks also appeared to be more diverse in Layout 2 to 4 than Layout 1. The trend seemed to be consistent across the experience groups except Layout 4 in the more experienced group. Also, we noticed that the difference between the two groups tended to be smaller in Layout 2 to 4 compared to the difference in Layout 1. An implication of this might be that an increased level of document representation can reduce the gap between the experience groups for the range of words used in the tasks. However, no statistical significance was found among the layouts, between Layout 1 and the cumulate of Layout 2-4, nor between the experience groups.

As for the query length, the overall difference among the layouts appeared to be small. While the less experienced group tended to submit a longer query in Layout 2 to 4, an inverse relationship was found in the more experienced group. However, no statistical significance was found among the layouts, between Layout 1 and the cumulate of Layout 2-4, nor between the experience groups.

5.1.2 Browsing and click-through URLs

Table 2 shows the number of result pages viewed and the number of click-through URLs made during the tasks. There are two aspects of viewed result pages shown in Table 2. The Result pages (>10) shows the number of pages viewed after the initial result pages, and the Result pages (All) represents the total number of result pages viewed during the tasks.

As can be seen, participants appeared to view more additional pages in Layout 3 and 4 compared to Layout 1, but this was not necessarily the case in Layout 2. This trend seemed to be consistent across the experience groups. An implication of this might be that participants were more encouraged by the thumbnails than TRS to examine further retrieved records. However, no statistical significance was found among the layouts, between Layout 1 and the cumulate of Layout 2-4, nor between the experience groups. The total number of result pages viewed tended to be larger in Layout 2 to 4 than Layout 1 due to the greater number of queries submitted to the interfaces, and this also seemed to be consistent across the experience groups. While no statistical significance was found between the individual layouts for Result pages (All), the difference between Layout 1 and the cumulate of Layout 2-4 was found to be significant by the Mann-Whitney U Test ($Z = -2.041, p = .020$). We did not find statistical significance of the difference between the experience groups.

Table 1: Query formulation

Number of queries	L1	L2	L3	L4	L2-4
Less Exp.	5.2 (3.2)	8.1 (6.1)	6.5 (3.8)	8.2 (5.2)	7.6 (5.0)
More Exp.	6.1 (4.6)	8.8 (6.7)	8.3 (5.6)	7.0 (6.2)	8.1 (6.0)
All participants	5.6 (3.9)	8.5 (6.3)	7.4 (4.7)	7.6 (5.6)	**7.8** (5.5)
Unique words	L1	L2	L3	L4	L2-4
Less Exp.	6.5 (2.8)	10.2 (6.5)	9.5 (7.8)	8.1 (4.5)	9.2 (6.3)
More Exp.	8.8 (6.0)	9.1 (5.6)	10.3 (6.6)	7.8 (3.2)	9.1 (5.3)
All participants	7.7 (4.8)	9.7 (6.0)	9.9 (7.1)	8.0 (3.8)	9.2 (5.8)
Query length	L1	L2	L3	L4	L2-4
Less Exp.	3.1 (1.2)	3.6 (1.5)	4.3 (1.4)	3.3 (1.0)	3.7 (1.4)
More Exp.	4.0 (2.0)	3.6 (1.2)	3.4 (0.9)	3.6 (1.9)	3.5 (1.4)
All participants	3.6 (2.6)	3.6 (1.3)	3.8 (1.3)	3.4 (1.5)	3.6 (1.4)

n (Less and More Exp. in L2-4) = 36; n (All participants in L2-4) = 72; n (Less and More Exp. in the rest) = 12; n (All participants in the rest) = 24

As for the number of click-through URLs, participants appeared to make fewer click-throughs in Layout 2-4 than Layout 1 although the overall difference among the layouts was relatively small. No statistical significance was found among the layouts, between Layout 1 and the cumulate of Layout 2-4, nor between the experience groups.

We then measured the average number of clicks made per result page. The click per page ratio was 2.2, 1.6, 1.3, 1.6 for Layout 1, 2, 3, and 4, respectively. The ratio of the cumulate of Layout 2-4 was 1.5. While no statistical significance was found between the individual layouts, the difference between Layout 1 and the cumulate of Layout 2-4 was found to be significant by the Mann-Whitney U Test ($Z = -1.964, p = .025$). Given that the underlying search engine was identical across the layouts, this seems to suggests that participants made more relevance judgements based on the document surrogates in Layout 2 to 4, thus, they did not have to visit the retrieved pages as often as they did in Layout 1. Another factor for this lower click per page ratio in Layout 2-4 compared to Layout 1 was likely the number of queries submitted to the interface. As we discussed above, participants tended to submit more queries when additional representations were available in the search result presentation. After some iterations of query reformulation, participants could gain a better idea of the topic and search results, thus, the number of clicks made on individual result pages can be reduced.
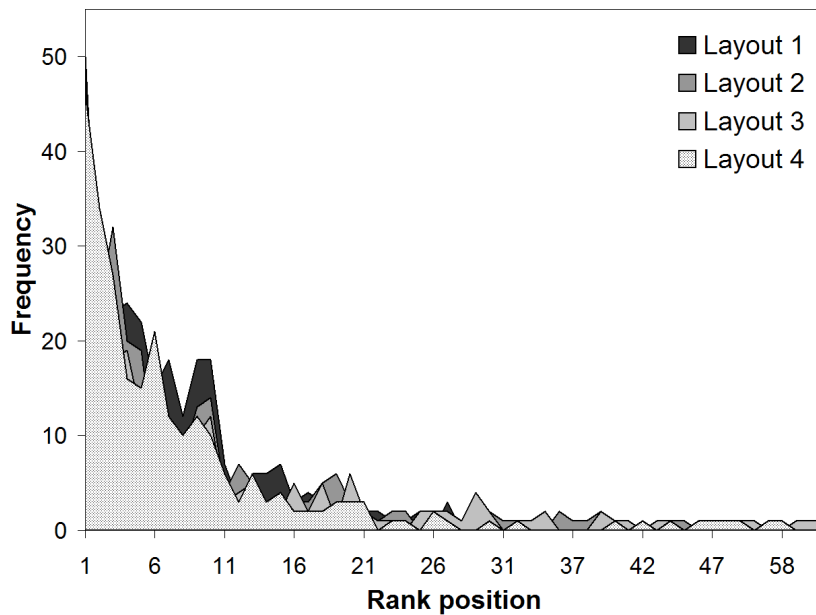
Figure 2 shows the distribution of click-through URLs in the four layouts. The X-axis

Table 2: Browsing and click-through URLs

Result pages (>10)	L1	L2	L3	L4	L2-4
Less Exp.	1.8 (2.0)	1.2 (1.6)	2.9 (2.6)	2.0 (2.3)	2.0 (2.3)
More Exp.	2.4 (3.5)	2.2 (2.3)	3.4 (4.1)	2.8 (1.4)	2.8 (2.8)
All participants	2.1 (2.8)	1.7 (2.0)	3.2 (3.4)	2.4 (1.9)	2.4 (2.6)
Result pages (All)	L1	L2	L3	L4	L2-4
Less Exp.	6.9 (3.5)	9.3 (6.4)	9.4 (3.7)	10.2 (6.6)	9.6 (5.6)
More Exp.	8.5 (7.3)	11.0 (6.6)	11.8 (7.1)	9.8 (5.9)	10.8 (6.4)
All participants	7.7 (5.7)	10.1 (6.4)	10.6 (5.7)	10.0 (6.1)	**10.2** (6.0)
Click-through	L1	L2	L3	L4	L2-4
Less Exp.	12.4 (5.8)	10.7 (5.8)	10.8 (4.2)	10.8 (4.2)	10.8 (4.7)
More Exp.	11.1 (10.0)	11.6 (6.0)	10.5 (6.9)	10.6 (6.6)	10.9 (6.3)
All participants	11.8 (8.0)	11.1 (5.8)	10.7 (5.6)	10.7 (4.2)	10.8 (5.5)

n (Less and More Exp. in L2-4) = 36; n (All participants in L2-4) = 72; n (Less and More Exp. in the rest) = 12; n (All participants in the rest) = 24

Figure 2: Rank positions of click-through pages (n = 1039).

represents the rank position of click-through URLs, and Y-axis represents the frequency of click-through. As can be seen, participants viewed and visited at a lower rank position more often in Layout 2 to 4 than Layout 1. For statistical test, we divided the rank positions into five bins (i.e., Rank at 1 to 20 went to Bin 1, 21-40 went to Bin 2, etc.) and run the Kruskal-Wallis Test on the data. The test shows that the difference between the four layouts was significant ($\chi^2(3) = 16.344$, $p = .001$). This appears to suggest that the additional representations can contribute to an increased level of exhaustivity in relevance assessments. Therefore, our observations made in this section, taken collectively, suggest that *an increased level of document representation can facilitate participants' browsing of the result pages.*

5.1.3 Task completion time

Table 3 shows the time taken to complete the tasks. Overall, the task completion time did not differ much over the layouts in both experience groups. However, the more experienced group tended to complete the tasks faster than the less experienced group. The difference between the two experience groups was found to be significant by the Mann-Whitney U Test ($Z = -1.861$, $p = .031$). However, no significance was found among the layouts, nor between Layout 1 and the cumulate of Layout 2-4. Therefore, an increased level of document representation was unlikely to have a significant effect on the task completion time.

Table 3: Task completion time

Time (min)	L1	L2	L3	L4	L2-4
Less Exp.	14.3 (1.5)	13.3 (2.1)	14.5 (1.4)	14.4 (2.0)	14.1 (1.9)
More Exp.	12.8 (3.5)	13.3 (3.7)	13.0 (3.0)	13.0 (3.2)	13.1 (3.1)
All participants	13.5 (2.8)	13.3 (2.9)	13.8 (2.5)	13.7 (2.5)	13.6 (2.6)

n (Less and More Exp. in L2-4) = 36; n (All participants in L2-4) = 72; n (Less and More Exp. in the rest) = 12; n (All participants in the rest) = 24

5.1.4 Summary

This section has examined the user interaction with the interfaces from three perspectives. The overall results discussed in this section suggest that an increased level of document representation available in the search results can facilitate user's interaction with the search interfaces. We have found that additional document representations can help searchers formulate more queries. There was also an indication that the additional representations can facilitate participants' browsing of the search results.

5.2 Participants' perceptions

This section presents the results of the participants' subjective assessments on relevance assessment, interfaces and layout features, and tasks they carried out. The assessments were indicated by a 7 point scale where a lower score represented a more positive assessment. In the tables, the mean value of the assessments is presented along with their standard deviation given in the brackets.

5.2.1 Perception on relevance assessment

Table 4 shows participants' assessments on the ease of relevance assessment during the tasks. The first question we asked was how easy it was to identify relevant documents in the search results, and the second question was how easy it was to identify *new* relevant information in the search results.

As can be seen, the less experienced group appeared to find Layout 2 to 4 easier to identify relevant documents compared to Layout 1. While no statistical significance was found between the individual layouts, the difference between Layout 1 and the cumulate of Layout 2-4 was found to be significant by the Mann-Whitney U Test ($Z = -1.819, p = .035$) in this group. This partially supports our finding discussed in Section 5.1.2 regarding the relevance assessments facilitated by the additional representations. On the other hand, the more experienced group appeared to find Layout 2 more

difficult to identify relevant documents. An implication of this might be that the more experienced searchers were accustomed to the existing layouts, thus, reluctant to use TRS in relevance assessments. However, no significance was found among the layouts, nor between Layout 1 and the cumulate of Layout 2-4 in this group.

As for the identification of new relevant information, the more experienced group appeared to find Layout 2 to 4 easier to carry out than Layout 1. The less experienced group appeared to find Layout 3 more difficult to do identify new information. However, no statistical significance was found among the layouts, nor between Layout 1 and the cumulate of Layout 2-4 in both groups.

5.2.2 Perception on the interface and layout features

Table 5 shows participants' assessments on the interfaces they used to carry out the tasks. The less experienced group appeared to find Layout 2 less useful to complete the tasks. On the other hand, the more experienced group appeared to find Layout 2 to 4 more useful than Layout 1 to complete the tasks. However, no statistical significance was found among the layouts, between Layout 1 and the cumulate of Layout 2-4, nor between the experience groups.

As for the effectiveness of the interfaces, the less experienced group appeared to find only Layout 4 more effective to complete the tasks compared to Layout 1. In both groups, Layout 2 tended to receive the least positive assessments among the layouts. An implication of this might be that the use of the visual form of additional representation can require less cognitive load than the textual form. However, no statistical significance was found among the layouts, between Layout 1 and the cumulate of Layout 2-4, nor

Table 4: Perception on the relevance assessment

Finding relevant	L1	L2	L3	L4	L2-4
Less Exp.	4.0 (1.9)	3.3 (2.1)	3.0 (1.7)	2.5 (1.7)	**2.9** (1.8)
More Exp.	3.3 (1.5)	3.9 (2.2)	2.7 (1.4)	2.5 (1.6)	3.0 (1.8)
All participants	3.6 (1.7)	3.6 (2.1)	2.8 (1.5)	2.5 (1.6)	3.0 (1.8)
Finding new	L1	L2	L3	L4	L2-4
Less Exp.	3.9 (1.7)	3.0 (1.2)	3.5 (1.6)	3.2 (1.2)	3.2 (1.3)
More Exp.	3.8 (1.5)	3.2 (1.6)	3.0 (1.2)	3.2 (0.9)	3.1 (1.3)
All participants	3.9 (1.6)	3.1 (1.4)	3.3 (1.4)	3.2 (1.0)	3.2 (1.3)

Range = 1-7; Lower = Easier; n (Less and More Exp. in L2-4) = 36; n (All participants in L2-4) = 72; n (Less and More Exp. in the rest) = 12; n (All participants in the rest) = 24

between the experience groups.

We then investigated the effectiveness of the individual layout features to gain further insight into the difference between the TRS and thumbnails. In our experiment, after each task, participants were asked to indicate to what extent each of the layout features contributed to their decisions of viewing documents from the search results. The result is shown in Table 6. The bottom row of Table 6 suggests that participants often found the title of retrieved documents the strongest factor in deciding which document to view from the search results, followed by the snippets. This seems to echo the finding of Tombros et al. (2005). Overall, participants rated the TRS as a stronger factor than the thumbnails in three tasks out of four. This suggests that the textual form is likely to have an effect in a wider context on user's relevance assessments than the visual form. Note, however, that the textual form might cause a higher level of cognitive load to the users although our data did not find the significant difference as shown in Table 5. When we compare TRS and thumbnail to Google's snippet, TRS was given a stronger score in Task 1 where participants were looking for the background information on student populations. On the other hand, the thumbnail was given a stronger score in Task 4 where participants were looking for key resources for designers handbags. An indication of this might be that the effectiveness of TRS and thumbnail can vary across the tasks. It also suggests that the benefits of TRS and thumbnails might be mutually exclusive. In other words, the thumbnails might be useful where TRS are found to be less effective, and vice versa.

We also measured the correlation of the layout features contribution. Table 7 shows Spearman correlation coefficient of seven layout features measured in our experiment.

Table 5: Perception on the interfaces

Interface usefulness	L1	L2	L3	L4	L2-4
Less Exp.	2.6 (1.2)	2.9 (1.4)	2.6 (0.6)	2.5 (1.6)	2.7 (1.2)
More Exp.	3.3 (1.8)	3.1 (1.5)	2.4 (0.9)	2.3 (0.9)	2.6 (1.2)
All participants	2.9 (1.5)	3.0 (1.4)	2.5 (0.8)	2.4 (1.2)	2.6 (1.2)
Interface effectiveness	L1	L2	L3	L4	L2-4
Less Exp.	3.8 (1.5)	4.0 (1.5)	3.9 (1.4)	3.1 (2.0)	3.7 (1.7)
More Exp.	3.7 (1.8)	4.4 (1.3)	3.1 (1.5)	3.3 (1.6)	3.6 (1.6)
All participants	3.7 (1.6)	4.2 (1.4)	3.5 (1.5)	3.2 (1.8)	3.6 (1.6)

Range = 1-7; Lower = Better; n (Less and More Exp. in L2-4) = 36; n (All participants in L2-4) = 72; n (Less and More Exp. in the rest) = 12; n (All participants in the rest) = 24

Table 6: Contribution of layout features

Task	Title	Snippet	TRS	Thumb.	URL	Size	Type
1	1.6 (1.1)	2.4 (1.7)	2.1 (1.2)	4.4 (2.2)	4.3 (1.9)	6.7 (0.8)	5.8 (1.7)
2	1.9 (1.5)	2.7 (1.8)	2.7 (1.4)	4.8 (2.0)	3.5 (2.3)	6.5 (1.0)	6.4 (1.2)
3	2.1 (1.6)	2.0 (1.3)	2.6 (1.7)	4.3 (1.8)	4.4 (2.0)	6.8 (0.5)	5.8 (1.7)
4	1.8 (1.2)	2.5 (1.5)	3.3 (1.7)	2.3 (1.5)	3.0 (2.0)	6.4 (1.3)	5.8 (1.7)
Total	1.8 (1.4)	2.4 (1.6)	2.6 (1.5)	4.0 (2.1)	3.8 (2.1)	6.6 (0.9)	5.9 (1.6)

Range = 1-7; Lower = Stronger; n (TRS and Thumb. in Task 1-4) = 12; n (the rest in Task 1-4) = 24; n (TRS and Thumb. in Total) = 48,; n (the rest in Total) = 96

The correlations that are statistically significant ($p \leq .05$) are highlighted in bold. As can be seen, TRS was found to have a positive correlation with both the title and snippet of the retrieved documents. On the other hand, the thumbnails had a small but significant negative correlation with the snippet and a positive correlation with URLs. The negative correlation with the snippet again suggests that the usefulness of additional textual representation can be mutually exclusive to the visual representation in user's relevance assessments on the search results. Also, thumbnails' positive correlation with URLs indicates that the visual information can be more influential when the genre or category of web pages is an important factor in the search tasks.

To investigate the latter observation, we conducted an analysis of the click-through pages across the tasks to examine the relationship between the effectiveness of TRS and thumbnails and high-level domains. Table 8 shows the distribution of click-through pages' domains across the tasks. As can be seen, while participants visited several types of domains in Task 1 and 3, over 95% of visits were concentrated in commercial domains in Task 2 and 4. This is interesting because participants indicated that thumbnails were least useful in Task 2 and most useful in Task 4 (see Table 7). This suggests that the domains of click-through pages can be an indication of the case where thumbnails might be helpful for searchers, but the domains alone will not be sufficient for the judgement.

Table 7: Correlation of layout features

	Title	Snippet	TRS	Thumb.	URL	Size	Type
TRS	**.410**	**.314**	1.000	-.175	-.202	.010	.147
Thumb.	.210	**-.265**	-.175	1.000	**.284**	**.247**	.051

n = 48

Table 8: Domains of click-through pages (Percentage)

Task	com	co.*	org	edu	ac.*	net	others
1	8.8	8.8	8.5	16.9	42.6	0.4	14.0
2	46.5	50.8	0.8	1.2	0.0	0.4	0.4
3	11.6	2.7	46.4	4.5	17.4	0.9	16.5
4	73.5	22.6	1.0	0.0	0.3	2.1	0.3

n (Task 1) = 272; n (Task 2) = 256; n (Task 3) = 224; n (Task 4) = 287

Table 9: Perception on the tasks

Task complexity	L1	L2	L3	L4	L2-4
Less Exp.	3.6 (2.3)	3.3 (2.0)	2.8 (1.9)	2.8 (1.8)	3.0 (1.9)
More Exp.	3.2 (1.6)	2.9 (1.4)	3.5 (2.0)	3.7 (1.7)	3.4 (1.7)
All participants	3.4 (2.0)	3.1 (1.7)	3.2 (1.9)	3.3 (1.8)	3.2 (1.8)
Task satisfaction	L1	L2	L3	L4	L2-4
Less Exp.	3.6 (1.6)	3.3 (2.1)	3.6 (2.1)	2.8 (1.5)	3.3 (1.9)
More Exp.	3.1 (1.6)	3.9 (1.9)	3.1 (1.5)	2.7 (2.0)	3.6 (1.6)
All participants	3.3 (1.6)	3.6 (2.0)	3.2 (2.0)	2.8 (1.7)	3.2 (1.9)

Range = 1-7; Lower (complexity) = Simpler; Lower (Satisfaction) = Better; n (Less and More Exp. in L2-4) = 36; n (All participants in L2-4) = 72; n (Less and More Exp. in the rest) = 12; n (All participants in the rest) = 24

5.2.3 Perception on the tasks

The last aspect of participants' perceptions is on the tasks they carried out. Table 9 shows participants' perceptions on the task complexity and satisfaction of task outcomes. As can be seen, the less experienced group tended to find the tasks simpler in Layout 2 to 4 than Layout 1. On the other hand, the more experienced group appeared to find the tasks more complex in Layout 3 and 4. An implication of this might be that the visual representation had some effect on the perception of task complexity in the more experienced group. However, no statistical significance was found among the layouts, between Layout 1 and the cumulate of Layout 2-4, nor between the experience groups.

As for the satisfaction of task outcomes, the less experienced group tended to be more satisfied with the outcomes in Layout 2 and 4 compared to Layout 1. On the other hand, only Layout 4 was given a more positive score than Layout 1 in the more experienced group. An implication of this might be that both textual and visual representations were necessary to achieve a greater level of satisfaction than Layout 1. However,

no statistical significance was found among the layouts, between Layout 1 and the cumulate of Layout 2-4, nor between the experience groups.

5.2.4 Summary

In this section we have examined participants' subjective assessments on the relevance assessments, interfaces, and tasks. We have found that the less experienced group tended to find it easier to identify relevant pages when the additional representations were available in the document surrogates. This partially supported the earlier discussion on the effect of additional representations in facilitating the browsing of the search results. The results also suggested that, in terms of the relevance assessments on the search results, the textual and visual representations were likely to have different contexts to contribute, and more importantly, these contexts can be mutually exclusive. On the other hand, we did not find a significant effect of the additional representations on participants' perceptions on task complexity and satisfaction.

5.3 System preference and participants' feedback

The last part of the result is based on participants' preference and feedback on the four layouts. Upon the completion of four tasks, participants were asked to rank the four layouts in the order of their preference. They were encouraged to rank the layouts based on the search experience of the given tasks. The most preferred layout was given Score 1 and the least was given Score 4. The counts of participants' ranking are presented in Table 10. As can be seen, the average rank (the bottom row) of Layout 2 and 4 was found to be comparable, followed by Layout 3. We ran the Friedman Test for the difference of system preference. The difference among the layouts were not found to be statistically significant. However, 20 out of 24 participants preferred the layouts with additional representations over Layout 1. 18 participants ranked Layout 2 over Layout 1, and 13 participants ranked Layout 3 or 4 over Layout 1. This suggests that participants did find cases where the additional representations were beneficial to some aspects of their search. The rest of this section presents participants' other feedback on the layouts that were captured during the exit interview. The codes T1 to T24 represent the ID of participants.

The reasons frequently mentioned about the advantage of Layout 1 (i.e., Baseline Google layout) was that it was simple, clean, and familiar (T3, T5, T8, T9, T11, T14, T17, T18). Perhaps, it is not surprising that they feel comfortable with the layout which they frequently use. However, several participants also commented that sometimes they did

Table 10: System preference

Preference	Layout 1	Layout 2	Layout 3	Layout 4
1 (Most)	4	6	3	11
2	6	8	8	2
3	2	7	10	5
4 (Least)	12	3	3	6
Average rank	2.92 (1.2)	2.29 (1.0)	2.54 (0.9)	2.25 (1.1)

n = 24

not get enough information from Layout 1 (T1, T5, T8, T20).

A positive aspect of Layout 2 frequently mentioned by participants was that they found additional textual information about retrieved documents useful in general (T6, T8, T19, T23, T24). Several participants also commented that TRS provided more contexts of query terms when the snippet was less informative (T5, T11, T14, T19, T20). Another comment was that TRS was useful when they were looking for specific or detailed information in the documents (T1, T4, T8, T16). This appears to explain a positive correlation with the titles and snippets discussed in Section 5.2.2. One commented that TRS was sometimes useful for finding related terms (T23).

A negative aspect of Layout 2 frequently mentioned was that it took up more space on screen (T1, T3, T5, T9, T14). A couple of participants commented that digesting the information provided by TRS sometimes caused them to stay in the result pages longer, as opposed to the retrieved pages (T10, T24). Similarly, a couple of participants mentioned that they preferred to visit the retrieved pages to get more information (T10, T22). This seems to correlate with our observation of the different level of cognitive load required in the textual and visual form of representation discussed in Section 5.2.2.

As for the advantages of Layout 3, participants commented that the layout was clean (T8, T22) and it was easier to convey the information from the thumbnails compared to TRS (T15, T18, T22). Several participants also commented that thumbnails helped them avoid to re-visit the same URLs or pages from the same site (T2, T12, T20, T22). This appears to echo our finding of facilitated relevance assessments discussed in Section 5.1.2. Another comment was that the thumbnails gave them an indication of web site genres (T8, T12, T18, T22). As can be seen, the information participants conveyed from the thumbnails appeared to vary. For example, one participant commented that the thumbnails were useful to guess the main target audience of sites (T5). Another participant (T10) stated that "the thumbnails let me know what kind of company/site it was (e.g., whether it was routing me to other companies or if it was a warehouse type dealer or a

small designer shop)".

A common complaint about the thumbnails was that sometimes they were too small to use and it would have been more useful if they were bigger (T4, T8, T14, T15, T19, T20). A couple of participants mentioned that the thumbnails were sometimes too abstract to use for a given request (T17) and less effective to get the detail of web pages (T3, T11). Most participants commented that this was the first time to use thumbnails in a search task. Therefore, it is not surprising that some participants found fewer appropriate contexts for the use of the thumbnails than TRS. It would be interesting to investigate the effect of TRS and thumbnails over the time in a longitude study.

The positive and negative comments reported above also applied to Layout 4. However, some feedback appeared to be unique to this layout. For example, participants commented that Layout 4 gave them a control on what information to use (T1, T13), and that it gave them more confident about documents' contents before visiting the pages (T4, T15). The most common complaint about Layout 4 was that the result page sometimes looked cluttered with too much information (T8, T9, T11, T14, T18).

6   Conclusive discussion

In this article, we presented a comparative evaluation of the textual and visual forms of document representation as additional components of document surrogates. A sentence-based summarisation technique was used to create the textual form of additional representation while the thumbnail image of web pages was used for the visual form of representation. Four layouts of search results were devised and evaluated in the experiment. Twenty-four people participated in our user study and each carried out four search tasks using the proposed layouts. The user interaction with the search interface was recorded during the experiment, participants' subjective assessments were established by a set of questionnaires, and other feedback was captured at the exit interview.

The implications for the design and development of search interfaces based on the findings of our study are as follows.

- *An increased level of document representation available in the search results is likely to help searchers formulate more queries.* The users of search engines are often reluctant to offer their effort in search (Jansen et al., 1998). Formulating queries that are well designed for IR systems is known as a challenge for many searchers (Baeza-Yates and Ribeiro-Neto, 1999). A number of approaches have been proposed to support user's query formulation (Efthimiadis, 1996). The results of our study indicate that one way to supplementary facilitate user's query formulation can be

to increase the level of document representation in the search result presentation. As such, our results also empirically support the advantage of poly-representation in interactive IR applications (Ingwersen, 1992).

- *Careful considerations are needed to determine the contexts where additional representations can be beneficial to searchers.* Overall, our results indicate that participants found cases where the additional representations were beneficial to some aspects of their search. However, we found no clear evidence that suggested either of the textual or visual forms of representations were consistently useful across the search tasks. Woodruff et al. (2002) suggested that the effectiveness of thumbnails can vary across the types of search tasks. The comparison of the thumbnails and TRS conducted in our study suggests that the textual information is likely to have a wider context to help searcher's relevance judgements compared to the visual information. More importantly, however, our results indicate that the contexts, where the two forms of representation are helpful, can be mutually exclusive.

- *The perception of additional representations can vary across users' search experience.* It has been suggested that the perception of the search interface features can vary across the level of search experience of the users (e.g., White et al., 2003a). The experienced searchers, who are used to control the search strategy with an existing interface (e.g., Layout 1), might be more reluctant to change their searching style for a new interface feature. In this study, we did not find the significant difference between the two experience groups regarding the level of interactions. However, our results indicate that less experienced searchers are more likely to find additional representations useful in relevance assessments.

- *A further insight into the effect of search result presentations can be gained in a holistic laboratory environment.* As discussed in Section 2, some of the existing work on the presentation of search results have involved participants in limited aspects of search (e.g., browsing) in the evaluation. Our results indicate that a subsequent study based on a holistic search environment, where participants are involved in all aspects of interactions, can provide further insight into the effects of search result presentation. The first point discussed in this section is just one such example.

One of our conclusions, therefore, might be that it is *safer* to show both of the textual and visual forms of additional representation in the search results. It might offer some searchers a greater degree of control in the selection of useful information to carry out searches. However, user's cognitive load is likely to increase when more elements are added to the search interface. Therefore, we argue that the search interface should be able to offer the right form of additional document representation in an appropriate

context or task. We carried out a preliminary study to investigate the association between the domains of click-through URLs and effectiveness of the thumbnails, but the result was inconclusive. Consequently, this study calls for further research on the understanding of users' search contexts and adaptive technique to capture their needs in an appropriate context. This study presented the cases where such advance can be used to facilitate some aspects of search experience.

It should be noted that this study was conducted on the live Internet using a single search engine, and no measure was taken to guarantee that the identical set of URLs was retrieved from the search engine in response to the same query. The entire experiment was carried out in the period of two weeks. Another limitation is the number of topics. While we aimed to test the effectiveness of the four layouts based on different types of tasks, the range of topics used was relatively small compared to the TREC Ad-Hoc experiments. Finally, although we designed our experiment in a more holistic search environment than some of existing work, this study was still carried out in a laboratory environment. Therefore, the findings of our study should be seen as exploratory and the generalisation of our implications can be limited.

7 Acknowledgements

References

Baeza-Yates, R. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. ACM Press, New York.

Beaulieu, M. (1997). Experiments on interfaces to support query expansion. *Journal of Documentation*, 53(1):8–19.

Beaulieu, M. and Gatford, M. J. (1998). Interactive okapi at trec-6. In Voorhees, E. M. and Harman, D. K., editors, *NIST Special Publication 500-240: The Sixth Text REtrieval Conference (TREC 6)*, pages 143–167, Gaithersburg, MD. NIST.

Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 56(1):71–90.

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

Chen, H. and Dumais, S. (2000). Bringing order to the web: Automatically categorizing search results. In *Proceedings of the CHI 2000 Conference on Human factors in computing systems*, pages 145–152, The Hague Netherlands. ACM.

Dziadosz, S. and Chandrasekar, R. (2002). Do thumbnail previews help users make better relevance decisions about web search results? In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 365–366, Tampere, Finland. ACM.

Efthimiadis, E. N. (1996). Query expansion. *Annual Review of Information Systems and Technology*, 31:121–187.

Greene, S., Marchionini, G., Plaisant, C., and Shneiderman, B. (2000). Previews and overviews in digital libraries: Designing surrogates to support visual information seeking. *Journal of the American Society for Information Science*, 51(4):380–393.

Gupta, A. and Jain, R. (1997). Visual information retrieval. *Communications of the ACM*, 40(5):70–79.

Hearst, M. A. and Pederson, J. O. (1996). Re-examining the cluster hypothesis: Scatter/gather on retrieval results. In Frei, H.-P., Harman, D., Schable, P., and Wilkinson, R., editors, *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 76–84, Zurich, Switzerland. ACM.

Ingwersen, P. (1992). *Information Retrieval Interaction*. Taylor Graham Publishing, London.

Jansen, B. J., Spink, A., Bateman, J., and Saracevic, T. (1998). 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):5–17.

Kelly, D., Deepak Dollu, V., and Fu, X. (2005). The loquacious user: a document-independent source of terms for query expansion. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 457–464, Salvador, Brazil. ACM Press.

Lam, H. and Baudiscch, P. (2005). Summary thumbnails: Readable overviews for small screen web browsers. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 681–690. ACM Press.

Larson, B. (2005). Practical implications of handling multiple contexts in the principle of polyrepresentation. In *Proceedings of the 5th International Conference on Conceptions of Library and Information Science (CoLIS 2005)*, Lecture Note on Computer Science (3507), pages 20–31, Glasgow, UK. Springer.

Robertson, S. E., Walker, S., Beaulieu, M. M., Gatford, M., and Payne, A. (1995). Okapi at trec-4. In Harman, D. K., editor, *NIST Special Publication 500-236: The Fourth Text REtrieval Conference (TREC-4)*, pages 73–97, Gaithersburg, MD. NIST.

Salton, G. (1971). *The SMART Retrieval System*. Prentice-Hall, Englewood Cliffs, NJ.

Shiri, A. A. and Revie, C. (2003). The effects of topic complexity and familiarity on cognitive and physical moves in a thesaurus-enhanced search environment. *Journal of Information Science*, 29(6):517–526.

Tague-Sutcliffe, J. (1992). The pragmatics of information retrieval experimentation, revisited. *Information Processing & Management*, 28(4):467–490.

Tombros, A., Ruthven, I., and Jose, J. M. (2005). How users assess web pages for information seeking. *Journal of the American Society for Information Science and Technology*, 56(4):327–344.

Tombros, A. and Sanderson, M. (1998). Advantages of query-biased summaries in information retrieval. In Croft, B. W., Moffat, A., and van Rijsbergen, C. J., editors, *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2–10, Melbourne, Australia. ACM.

White, R. (2004). *Implicit Feedback for Interactive Information Retrieval*. Phd thesis, Department of Computing Science, University of Glasgow, Glasgow, UK.

White, R., Jose, J. M., and Ruthven, I. (2003a). A task oriented-study on the influencing effects of query-biased summarisation in web searching. *Information Processing & Management*, 9(5):707–733.

White, R. W., Jose, J. M., and Ruthven, I. (2003b). A granular approach to web search result presentation. In *Proceedings of the Ninth IFIP TC13 International Conference on Human-Computer Interaction (INTERACT 2003)*, pages 220–227, Zürich, Switzerland. IOS Press.

Woodruff, A., Rosenholtz, R., Morrison, J. B., Faulring, A., and Pirolli, P. (2002). A comparison of the use of text summaries, plain thumbnails, and enhanced thumbnails for web search tasks. *Journal of the American Society for Information Science and Technology*, 53(2):172 – 185.

Zamir, O. and Etzioni, O. (1999). Grouper: A dynamic clustering interface to web search results. *Computer Networks*, 31:11–16.