Using Top-Ranking Sentences to Facilitate Effective Information Access

Ryen W. White
Institute of Advanced Computer Studies, University of Maryland, College Park, MD 20742. E-mail:
ryen@umiacs.umd.edu

Joemon M. Jose
Department of Computing Science, University of Glasgow, Glasgow, Scotland G12 8RZ.
E-mail: jj@dcs.gla.ac.uk

Ian Ruthven
Department of Computer and Information Sciences, University of Strathclyde, Glasgow, Scotland G1 1XH.
E-mail: ir@cis.strath.ac.uk

Web searchers typically fail to view search results beyond the first page nor fully examine those results presented to them. In this article we describe an approach that encourages a deeper examination of the contents of the document set retrieved in response to a searcher's query. The approach shifts the focus of perusal and interaction away from potentially uninformative document surrogates (such as titles, sentence fragments, and URLs) to actual document content, and uses this content to drive the information seeking process. Current search interfaces assume searchers examine results document-by-document. In contrast our approach extracts, ranks, and presents the contents of the top-ranked document set. We use query-relevant top-ranking sentences extracted from the top documents at retrieval time as fine-grained representations of top-ranked document content and, when combined in a ranked list, an overview of these documents. The interaction of the searcher provides implicit evidence that is used to reorder the sentences where appropriate. We evaluate our approach in three separate user studies, each applying these sentences in a different way. The findings of these studies show that top-ranking sentences can facilitate effective information access.

Introduction

The value of systems that help Web searchers find relevant information is becoming increasingly apparent. Such systems involve a searcher, with a need for information,

motivated by a gap in their current state of knowledge (Belkin, Oddy, & Brooks, 1982), seeking the information required to close the gap, solve the problem that initiated the seeking, and satisfy their need. Typically, searchers are expected to express this need via a set of query terms submitted to the search system. This query is compared to each document in the collection, and a set of potentially relevant documents is returned. These may not be completely relevant, and it is the relevant (or partially relevant) parts that contribute most to satisfying information needs.

Ranking whole documents assumes that all of a document conforms to relevance/matching criteria. Research into summarization (Amitay & Paris, 2000; Tombros & Sanderson, 1998) and visualization (Dziadosz & Chandrasekar, 2002; Hearst, 1995) have tackled this problem, but still return document lists to searchers. Other representations of search results have been tested. These either present the searcher with an unfamiliar, usually graphical interface that imposes an increased cognitive burden (Ahlberg & Shneiderman, 1994; Hemmje, 1995), or consider documents as the finest level of granularity for result presentation (Chen & Dumais, 2000; Dumais, Cutrell, & Chen, 2001).

The transformation of an information need into a search expression, or query, can be a cognitively expensive and demanding process (Goecks & Shavlik, 2000). This is typically regarded as one of the most challenging activities in information seeking (Cool et al., 1996). However, a searcher may face even more difficulty when interpreting and assessing the relevance of the returned documents. Searchers are typically unwilling to visit individual documents to gauge relevance and base judgments on document *surrogates*, such as titles, abstracts, and URLs, presented by the retrieval system.

Information retrieval (IR) systems were originally devised for the retrieval of documents from homogeneous corpora, such as newspaper collections or library index cards. Document surrogates were usually created by experts, such as librarians or professional cataloguers. However, the growth in size, dynamism, and heterogeneity of these collections necessitated the development of automated indexing techniques that led to a reduction in the quality of the surrogates created that was documented as early as the mid-1960s (Edmundson, 1964).

Presenting lists of document surrogates has remained a popular method of presenting search results. While conveniently packaging information and providing a ranking based on estimated utility, such lists can also be restrictive. They encourage searchers to read, interpret, and assess documents and their surrogates individually. It may be the information in the document, complemented by the document surrogates, which searchers require to close the knowledge gap that drives their seeking. These automatically generated surrogates are an intermediate step between the submission of a query and the perusal of one or more documents returned in response to that query. However, the indicative worth of these surrogates has been shown to be questionable, which can make the assessment of document relevance problematic (White, Jose, & Ruthven, 2003b).

In this article we describe and evaluate an approach that encourages a deeper examination of documents at the results interface and blurs interdocument boundaries. We shift the focus of interaction from document surrogates to document content, and rank this content regardless of its source. For this purpose we use top-ranking sentences (TRS) taken from the top-retrieved documents, ranked based on the query, and presented in a list to the searcher. These are the most potentially useful sentences in the top-ranked documents, extracted and scored according to factors such as their position in the source document (initial introductory sentences are preferred), the words they contain (those emphasized by the Web page author, e.g., emboldened terms, and words in the document title or document headings are preferred), and the proportion of query terms they contain. The latter component—scoring by query terms—ensures that the sentences extracted are query-relevant. Through presenting the sentences chosen from each document in a ranked list, ranked with respect to sentence score and independent of source document, we present a query-biased overview of the retrieved set's content. In this way, highly relevant content from lower ranking documents, that might not have been viewed, simply because of its resident document's rank position, is made accessible to the searcher. Figure 1 shows part of a list of top-ranking sentences taken from one of the three studies described in this article. The query terms "dust" and "allergies" are highlighted by the system.

Top-ranking sentences help searchers target potentially useful information. Potentially relevant sentences appear near the top of the list, guiding searchers towards the answer they seek or documents of interest. The sentences encourage interaction with the content of the retrieved document set, an

FIG. 1. An excerpt from a list of top-ranking sentences (query is "dust allergies").

approach we call *content-driven information seeking* (CDIS). This is in contrast to query-driven approaches, where searchers proactively seek information through the query they provide.

Typically, Web-search systems use lists of document surrogates to present their search results. This forces searchers to make two steps when assessing document relevance; first assess the surrogate, then perhaps peruse and assess the document (Paice, 1990). Such systems enforce a *pull* information seeking strategy, where searchers are proactive in locating potentially relevant information from within documents. In CDIS, it is the system that acts proactively, presenting the searcher with potentially relevant sentences taken from the document set at retrieval-time. The system uses a *push* approach, where potentially useful information is extracted from each document and proactively pushed to the searcher at the results interface. Searchers have to spend less time locating potentially useful information.

To illustrate the worth of our approach we describe three related user studies, using 58 different subjects and conducted over a period of 9 months. In the analysis of the findings we focus on the relationship between the studies and qualitative subject perceptions of the approaches we describe. Hereafter we refer to the three studies as *TRSPresentation*, *TRSFeedback*, and *TRSDocument*.[1] Due to variations in subjects, systems, and search tasks, it is difficult to make comparisons between the quantitative results obtained in each study. For this reason, quantitative results of the experiments are generally not presented in this article, only the subject perceptions of the techniques employed. The quantitative results for all three studies can be found in White, Jose, and Ruthven (TRSPresentation; 2003a), White, Ruthven, and Jose (TRSFeedback 2002b) and White, Ruthven, and Jose (TRSDocument; 2002a). This article describes how subjects use top-ranking sentence interfaces for their search, how this differs from traditional search methods, and the reason why top-ranking sentence interfaces are preferred over traditional forms of result presentation.

---

[1]TRSPresentation = top-ranking sentences for result presentation; TRSFeedback = top-ranking sentences for feedback decisions; TRSDocument = top-ranking sentences for document access.

The remainder of this article is structured as follows. In the next section we further describe the pull and push information seeking strategies. The User Studies section provides details on the user studies undertaken, including the experimental methodology employed, the systems created, and the relationship between the three studies. In the following section we describe the results and discuss their implications for the design of Web search interfaces; we present are conclusions in the last section.

Information Seeking Strategies

Searchers approach IR systems with a need for information. The information required to satisfy this need transcends document boundaries and is a culmination of the knowledge gleaned from documents examined during the search session (Belkin, 1984). However, returning a ranked list of documents does not fit well with this model. The list restricts the interaction and general information seeking behavior of searchers; they are forced to examine search results individually.

Most Web search interfaces present the searcher with little information with which to decide whether or not to view a retrieved document. Typically, the only information shown is the document title, URL, and short (one–two line) sentence fragments containing the query terms and a small number of fore and aft terms to afford context. These snippets are rarely full sentences, and are often separated by ellipses. Figure 2 shows an example of such surrogate information.

This information is important because searchers use it to make decisions about what documents to view (Furnas, 1997). In result lists, searchers assess document relevance externally, based on what they can infer from their surrogates. On the Web, authors assign document titles; the extent to which these titles are indicative of content can vary. This differs from the static homogeneous collections described earlier, where there is consistency in the titles/headlines assigned. To provide searchers with representations that are truly indicative, it is necessary to go deeper into the documents, extracting their content at a fine level of granularity but with increased contextual coherence (i.e., with whole sentences). Through using top-ranking sentences, IR systems can present the query terms in the local context in which they are used within retrieved pages and encourage interaction with results interfaces.

The relative success of IR systems can depend on at least two factors: (a) the question posed by the searcher, and (b) the searcher's ability to successfully interpret the response offered. If (a) and (b) are handled well then the probability of a successful search is increased; this scenario is often not realized. Studies have shown that searchers refrain from using the advanced search facilities that many Web search systems offer and display limited interaction with search engine interfaces (Crouch, Crouch, Chen, & Holtz, 2002; Jansen, Spink, & Saracevic, 2000). The approach described in this article encourages more interaction with search interfaces and in some cases, uses this interaction to make decisions on the searcher's behalf.

In this section two contrasting strategies for result presentation are described: One presents the searcher with surrogate document representations (e.g., titles, sentence fragments, and URLs) and relies thereafter on the searcher to visit the document, the other is a version of CDIS that proactively pushes potentially relevant content towards searchers. These two differing forms of result presentation encourage different information seeking strategies and different emphasis. The "need" in online searching is typically one for information. The perusal of ranked lists of documents may be an unnecessary step between query submission and direct access to this information. In what follows we describe what we refer to as *pull* and *push* information seeking, and the differences between the two approaches.

Pull Approach

In the pull approach the searcher must be proactive. They assess the value of documents externally based on document surrogates such as titles, sentence fragments, and URLs, and examine search results on a document-by-document basis. The document is considered as the finest level of granularity and the system presents a ranked list of documents based on the estimated utility of each in relation to the searcher's submitted query.

The sentence fragments may provide the motivation with which to visit a document; however, once inside the document the searcher has to locate the information then gauge its relevance in the context. Saracevic (1975) proposed, that as searchers move through the various stages of their information need evolution, where their need potentially becomes more certain (Ingwersen, 1994), their judgments of relevance are likely to change to take into account their newly encountered knowledge. Documents that are relevant at the start of the search may not be at the close. They are potentially cumbersome entities that can be completely, partially, or not relevant. It may not be prudent for a searcher to spend much time reading a document to assess whether the document is relevant, and it may simply not be possible to assess a document's relevance in a short time.

In the pull approach, the searcher is responsible for formulating the initial query and for further revising this query as their search proceeds. They are burdened with the responsibility to select additional query words and drive their own search. This can be problematic if the information need is vague (Spink, Griesdorf, & Bateman, 1998) or searchers are unfamiliar with the collection being searched of the retrieval environment (Salton & Buckley, 1990).

House **Dust** Allergy
... Advice From Your Allergist – House **Dust** Allergy. House **dust** allergy is common even in clean homes. ... Why does house **dust** cause allergic reactions? ...
allergy.mcg.edu/advice/dust.html - 16k - Cached - Similar pages

FIG. 2.   Web search engine abstract for the query "dust allergies."

Push Approach

In the push approach, the search system acts proactively, presents information extracted from the retrieved documents at query-time, and restructures this information based on implicit feedback. This is an implementation of the CDIS approach described earlier. Searchers can use the content presented to them at the interface (e.g., the top-ranking sentences) to guide them through their search. These sentences shift the emphasis from retrieved documents to the content of those documents. Ranking sentences in this way provides searchers with a query-relevant overview of retrieved documents. The focus of perusal and interaction is no longer a ranked list of document surrogates offering an external view of documents. Searcher attention is instead focused on potentially useful parts of retrieved documents. Therefore, less time need be spent locating useful information, meaning more time can be spent assessing its value.

Although the Web is a heterogeneous collection, the documents that are returned in response to a query, however different in nature and characteristics, are topically related to the query in some way. Presenting a list of top-ranking sentences not only provides a ranking with respect to the searcher's information need, but allows sentences to be seen in relation to other sentences from other documents. Therefore, searchers can view a ranking of content, not document lists that appear to be ranked based on content.

As well as using the top-ranking sentences to convey potentially relevant information, the sentences can also be reordered to communicate changes in the search system's formulation of relevance. Implicit feedback systems make inferences of what is relevant based on searcher interaction. They do not intrude on the searcher's primary line of activity (i.e., satisfying their information need). The treatment by the system of the searcher's action as evidence of relevance is secondary to the main task, which is to respond to the searcher's instruction (Furner, 2002). In traditional relevance feedback systems, the function of making judgments is intentional, and specifically for the purpose of helping the system build up a richer body of evidence on what information is relevant. The ultimate goal of information seeking is to satisfy an information need, not to rate documents. Systems that use implicit feedback to predict information needs and enhance search queries fit better with this goal. Implicit feedback can be seen as an enabling technique for the push paradigm and in the next section we describe how this feedback is captured.

Implicit Feedback

Implicit feedback systems typically use measures such as document reading time, scrolling, and interaction to make decisions on what information is relevant (Claypool, Waseda, & Brown, 2000). However, these systems typically assume that searchers will view and interact with relevant documents more than nonrelevant documents. These assumptions are context-dependent and can vary greatly between searchers. The approach used for implicit feedback in this study makes a potentially more robust assumption; searchers will try to view relevant information. Through monitoring the information searchers interact with, search systems can approximate search interests. This is made possible because the interface components the search interfaces present are smaller than the full-text of documents, allowing relevance information to be communicated more accurately.

In TRSFeedback and TRSDocument some of the experimental systems use evidence gathered via implicit feedback to restructure the retrieved information during the search. In these systems, each retrieved document has an associated summary composed of the best four top-ranking sentences that appear on the interface at the searcher's request. The viewing of this summary is regarded as an indication of interest in the information it contains and is used as an indication of searcher interests.

These relevance indications are used by the systems to reorder the top-ranking sentences. Sentences are small and the differences in sentence scores between sentences are also small. Should there be a slight change in the system's formulation of the information need, a list of sentences is much more likely to change than, say, a list of documents. At no point, in any experimental system, is the searcher shown the expanded query; they are only shown the *effect* of the query (i.e., the reordered top-ranking sentence list). Reordering the sentence list based on implicit feedback means it represents the system's current formulation of the searcher's information need. As this formulation is based solely on the viewed information, the system is able to form reasonable approximations on what information is relevant. As the searcher becomes more sure of their need, or indeed as the need changes, the search system can adapt, select new query terms, and use this query to update the ordering of the top-ranking sentences list to reflect this change.

The user studies described herein present subjects with search interfaces that may be unfamiliar to them. During these studies we felt that it was not necessary for subjects to see the contents of the modified query to use these interfaces effectively. This was the case, but some experimental subjects suggested that they may feel more comfortable with using the interfaces if they could view and manipulate the revised query.

Comparison of Push and Pull Approaches

The push approach extracts and presents potentially useful information to the searcher at the results interface. This content discourages searchers from examining documents individually and encourages the assessment of information resident in the result set regardless of its resident document. In contrast, the pull approach encourages searchers to assess documents externally, basing relevance assessments on the information presented in result lists.

In the push approach, sentences from documents are extracted in real-time and shown to the searcher at the results interface. In contrast, the pull approach provides less information to the searcher and they see only an external view of

TABLE 1. Differences between push and pull approaches.

	Approach	
Factor	Push	Pull
Information extraction	System	Searcher
Finest granularity	Sentence	Document
Results perusal	Sentence/scanning sentences	Document-by-document
Facilitates interaction	Sentence (content)	Surrogate
Assess document value	Internally	Externally
System formulation of information needs	Static/dynamic	Static

the document. To find relevant information, they must first visit, then locate information inside documents. The differences between the approaches are mainly in the nature of search activity and how information is presented at the search interface. Table 1 shows other differences between the two approaches.

As Table 1 shows, the push approach uses smaller document representations, allows searchers to assess the value of information from within documents and adapts its formulation of information needs dynamically, without searcher instruction. It is only in push systems that do not use implicit feedback techniques where the system's internal queries are static until the next searcher-initiated query iteration. The push approach selects and presents potentially relevant sentences at the results interface; visiting documents a secondary activity, and the required information may be found directly at the results interface. In the pull approach, visiting documents is the main search activity, and unless the task is trivial, searchers will have to visit documents to find relevant information.

In the next section we describe a series of related user studies that test the worth of the content-driven information seeking approach using top-ranking sentences. These preliminary studies show that these techniques can be effective and are liked by searchers.

User Studies

Three user studies tested the worth of top-ranking sentences in different information seeking contexts. The results from these studies are summarized in this article, each of which uses these sentences in a different way. In the TRSPresentation study the ranked sentences are used as an alternative to document lists, shifting searcher attention from the document surrogates to the document content. TRSFeedback uses the sentences to reflect the use of two contrasting relevance feedback techniques. Finally, TRSDocument uses the sentences to encourage interaction with the retrieved set, to reflect change in the formulations of information needs and to complement, rather than replace, document lists. Each study involved real searchers and different types of information seeking scenario. The experimental systems selected top-ranking sentences in real-time, when the query

was submitted. This had the potential to cause delays in system operation.[2] Below the generic experimental methodology is described, as are the experimental interfaces used, the tasks assigned and the relationship between studies.

Experimental Methodology

In all three studies human subjects were recruited from a variety of backgrounds and assigned realistic search scenarios. The length of the experiment varied between 60–90 minutes depending on the number of experimental systems. The studies followed a common experimental procedure:

1. Introductory orientation
2. Presearch/demographic questionnaire
3. For each system in the study:
   (a) Short training session
   (b) Distribute search scenario and give subjects an opportunity to clarify any ambiguities
   (c) 10–15 minutes allowed for searcher to attempt the task
   (d) A postsearch questionnaire
4. A final questionnaire
5. An informal discussion (optional)[3]

The different experimental hypotheses between studies necessitated minor differences in the experimental methodology used.

Subjects

The recruitment of experimental subjects in these studies was specifically aimed at targeting two groups of subjects: inexperienced and experienced. Two of the three studies (TRSPresentation and TRSDocument) classified subjects in this way. In these studies the classification was made based on subjects' responses on questions about their experience and their own opinion of their skill level. TRSFeedback did not classify subjects. The number of subjects in each study varied between 16 and 24, the majority of whom were university students. All studies use a within-subjects experimental design meaning that subjects used all experimental systems. Latin and Graeco–Latin squares (Tague-Sutcliffe, 1992) are used to control subjects' learning effects between tasks and systems where appropriate.

Tasks

In TRSPresentation and TRSDocument subjects attempted combinations of tasks from the following categories: fact search (e.g., finding a named person's current e-mail address), decision search (e.g., choosing the best impressionist art

---

[2]In each study top-ranking sentences were taken from only the top 30 documents to ensure the systems responded in a timely manner.

[3]The informal discussion was initiated at the subject's or experimenter's request. An opportunity to take part in such a discussion was offered to all participants.

museum) and background search (e.g., finding information on dust allergies) (White et al., 2002a). Each search task was placed within a simulated work task situation (Borlund, 2000), which created realistic search scenarios and allowed personal assessments of what information was relevant. TRS-Feedback was carried out as part of the 2001 Task Retrieval Conference (TREC 2001 Interactive Track; Hersh & Over, 2001). The tasks were assigned by the track and divided up into four categories: Medical, Buying, Travel, and Project. Subjects attempted a task from each category.

Interfaces

Each of the three studies used top-ranking sentences to facilitate information access, encourage interaction and convey system decisions. In this section we describe the experimental interfaces used in each of these studies and explain the role of the top-ranking sentences in each interface.

TRSPresentation study. This study investigates the effectiveness of presenting a list of top-ranking sentences rather than a list of documents. The top-ranking sentences approach is compared against two interfaces that use traditional result presentation techniques (i.e., a list of document surrogates). One experimental system directly presents the results from the underlying search engines and the other uses the top-ranking sentences as document summaries, presented below the document title. The experimental interface, shown in Figure 3, consists of two main components; the top-ranking sentences (that replace the traditional document list) and a document pop-up window, that shows the searcher more information about a particular document.

Initially there is no direct association between a top-ranking sentence and its source document, i.e., there is no indication to the searcher of which document supplied each

sentence. To view the association, the searcher must move the mouse pointer over a sentence. When this occurs, the sentence is highlighted and a window pops up next to it. Displaying this window next to the sentence, instead of in a fixed position on the screen, makes the sentence–document relationship more lucid. In the window the searcher is shown the document title, URL, and the rank position and content of any other sentences from that document that occur in the list of top-ranking sentences. If no other sentences appear, an appropriate message is shown. To visit a document the searcher must click the highlighted sentence, or any sentences in the pop-up window. In this interface, the sentences encourage searchers to examine search results more deeply (through their content) and broadly (across a greater number of search results).

TRSFeedback study. In this study we tested the substitutability of implicit feedback for explicit feedback in Web retrieval. For this purpose we developed two interfaces, one where the system endeavours to estimate relevance by mining the searcher's interaction (implicit system) and one where subjects had to explicitly mark information as relevant (explicit system). In both systems the top-ranking sentences list updates in the presence of relevance information. They adapt to the context of the search by basing their term selection on relevance information provided during the examination of results. Techniques that are viable substitutes for explicit relevance feedback demonstrate the value of implicit feedback in information seeking environments. The assumption is that viewing a document summary is an indication of searcher interests. Figure 4 shows the interface to the implicit system.

After each relevance indication the summaries from the assessed relevant documents (explicit system) or assumed relevant documents (implicit system) are used to generate a list of possible query modification terms using the *wpq* formula (Robertson, 1990). The most useful modification terms
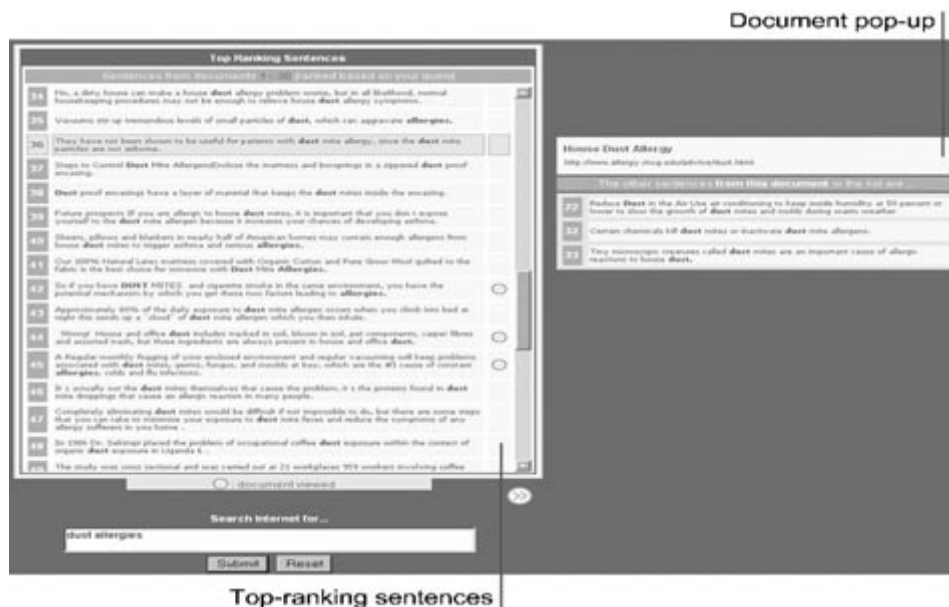

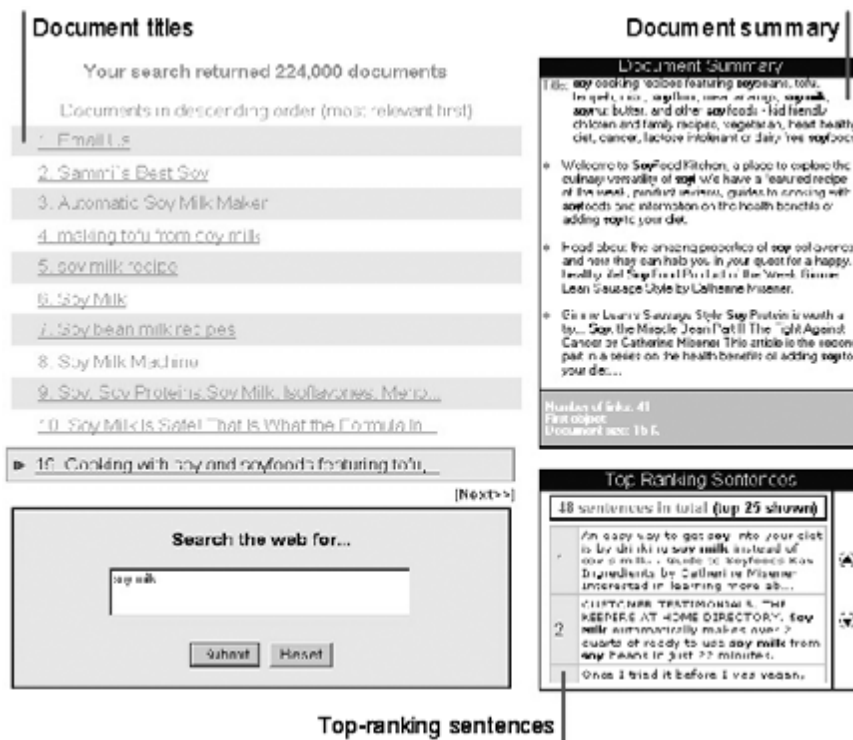FIG. 3.   The experimental interface for the TRSPresentation study.

FIG. 4.   The experimental interface for the TRSFeedback study.

are chosen from this list and added to the searcher's original query. These terms are chosen from all assumed relevant summaries (i.e., those viewed so far or those from documents they have checked), and used to reorder the list of top-ranking sentences. In this study the sentences are used to communicate the effects of relevance feedback decisions.

TRSDocument study.   In a similar way to TRSFeedback, the experimental interface in this study applied implicit feedback techniques on a list of top-ranking sentences. In this study, rather than communicating relevance feedback decisions, the sentences were used to facilitate access to retrieved documents. In this study, the experimental system is compared against a baseline summarization system used in White et al. (2003b) and a system where the order of the sentence list does not change over time and the query is assumed to be constant within an individual search iteration. Figure 5 shows the interface used in the static baseline system and the experimental system. The baseline summarization system used the same interface components other than the list of top-ranking sentences.

The experimental system uses implicit feedback given by a searcher while interacting with the interface. As in the implicit system in TRSFeedback, the implicit feedback is the evidence the searcher gives by viewing a document summary. The searcher does not give the information with the explicit purpose of changing the search results.

In TRSFeedback the system interprets every summary view as an indication of relevance. This led to problems of accidental "mouseover," with searchers passing over titles en route to those that interested them. In this study, the system implemented a timing mechanism that dealt with this problem and allowed us to base the implicit feedback on the length of time a searcher spent viewing a summary (White et al., 2002a). In this study we assume that summaries which searchers view for longer than expected are those that contain information similar to that desired by searchers. Viewing time was used as a determinant of whether a summary was relevant. Summaries were used since the systems can detect which summaries a searcher has assessed and for how long (unlike titles) and searchers tend to view more summaries than Web pages leading to more evidence for the techniques employed. Any summary that the system believes contains relevant information is used for query modification.

In the three systems in this study, one presented only titles and summaries, and two used titles, summaries, top-ranking sentences. From the latter interfaces, one reordered the sentences (in light of relevance information) and the other did not. In this study, the sentences were used to facilitate interaction with retrieved documents.

All interfaces presented in this section encourage a deeper examination of search results and some used implicit feedback techniques to adapt the display in light of searcher interaction. In the next section we describe the relationship between the interfaces and the user studies that test them.

Interstudy Relationship

The studies all used top-ranking sentences, but for a different purpose and to test different sets of hypotheses. Table 2 illustrates the main factors of each study.

TABLE 2. The main experimental factors in the three user studies.

	Study		
Factor	TRSPresentation	TRSFeedback	TRSDocument
Hypotheses	1. Top-ranking sentences as an alternative to a document abstract 2. Top-ranking sentences increases awareness of result set content and is preferred by users 3. Top-ranking sentences lead to improved perceptions of task success, actual task success and agreeability across all tasks	1. Implicit relevance feedback is a viable substitute for explicit relevance feedback in Web retrieval—tested via the reordering of the top-ranking sentences list	1. The use of top-ranking sentences encourages searchers to interact more fully with the retrieval results (i.e., documents) and whether this leads to more effective searching 2. Implicit feedback improves searchers' perceptions of the system and leads to more effective interaction
Measured	Search effectiveness, user perceptions	Search effectiveness, user perceptions	Search effectiveness, user perceptions
Number of systems	3	2	3
Systems (type)	1. Search engine baseline 2. TRS as abstracts 3. TRS as list	1. Implicit feedback 2. Explicit feedback	1. Summarization baseline 2. Summarization/TRS 3. Summarization/TRS and implicit feedback
Subjects	20	16	24
Grouping	10 Inexperienced 10 Experienced	None	12 Inexperienced 12 Experienced
Age	Average = 23.8 years Range = 32 years (17:49)	Average = 24.75 years Range = 11 years	Average = 24.73 years Range = 33 years (16:49)
Internet	Inexperienced = 4.2 h	14 h	Inexperienced = 4.1 h
Usage/week	Experienced = 32.6 h		Experienced = 29.8 h
Tasks	Three simulated work tasks (fact, decision and background)	Four each of Medical, Buying, Travel, and Project	Three simulated work tasks (fact, decision and background)
Experimental design	Graeco–Latin square	Latin square	Latin square
Tasks per user	3	4	3
Time per task	10 minutes	10 minutes	10 minutes
Data collection	Five questionnaires (1 demographic, 3 system and 1 final) Background logging	Five questionnaires (1 demographic and 4 system) Background logging	Five questionnaires (1 demographic, 3 system and 1 final) Background logging Semi-structured interviews

Note. TRSPresentation = top-ranking sentences for result presentation; TRSFeedback = top-ranking sentences for feedback decisions; TRSDocument = top-ranking sentences for document access; TRS = top-ranking sentences.
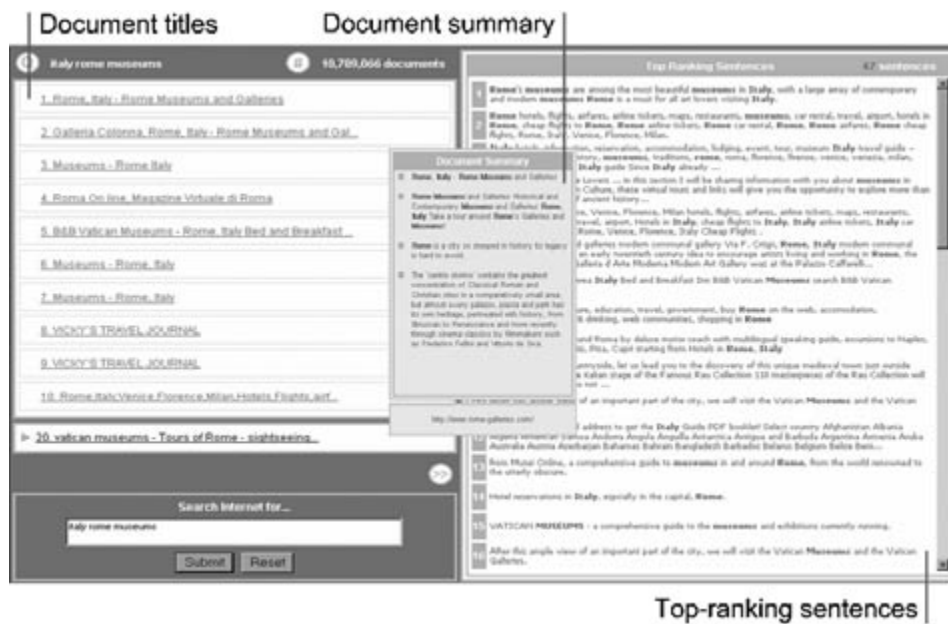

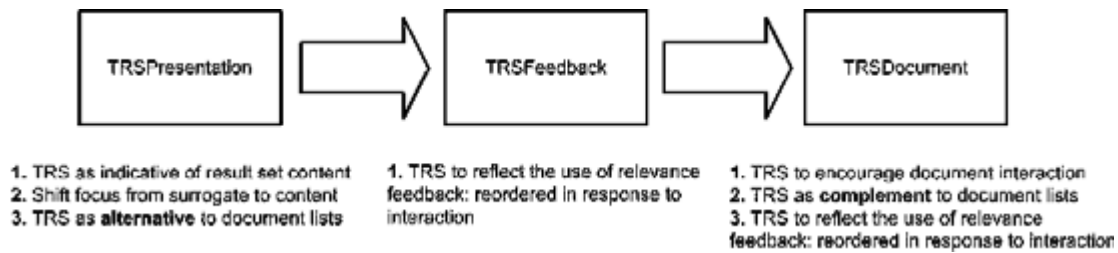FIG. 5. The experimental interface for the TRSDocument study.

FIG. 6. The relationship between the three user studies.

In TRSPresentation we encourage searchers to employ other ways of examining search results, and use the sentence list as a replacement for the document list. In TRSFeedback, top-ranking sentences were used to communicate system decisions in a comparison between implicit and explicit relevance feedback. TRSDocument uses the sentences to facilitate interaction with the top-ranked documents. The experimental system in TRSDocument still promotes the viewing of documents, but uses both documents and top-ranking sentences. The content still drives the interaction with documents via the query-relevant sentences they contain.

The three studies are related and illustrate the initial stages of the development of our techniques. Top-ranking sentences are first introduced as a replacement for document lists (TRSPresentation) then used to study the substitutability of implicit and explicit feedback (TRSFeedback). We finish by using both documents and sentences in a more intricate form of implicit feedback (TRSDocument), based on the proof of substitutability that TRSFeedback provided us with. Figure 6 shows the development between studies.

Top-ranking sentences drive searcher interaction. The same underlying motivation for their use applies in all three studies; ranking the content of the retrieved document set, rather than the documents themselves helps searchers. In the next section qualitative results from the studies are presented and the implications of them discussed.

Findings and Discussion

In this section we present and discuss the qualitative findings of the user studies. The quantitative results, and more system details, have already been presented in White et al. (TRSDocument, 2002a; TRSFeedback, 2002b; TRSPresentation, 2003a). Because the studies were conducted with different subjects, on different systems, at different times, direct comparisons across studies is difficult. Therefore, we focus mainly on subject opinions of the search process, the top-ranking sentences, and the implicit feedback used to reorder the sentences.

Search Process

Kuhlthau (1991) introduced a six-stage model of the information search process (ISP), where searchers seek meaning from information to enhance their knowledge of their current problem or search topic. In this section, where appropriate, we discuss the findings of the user studies in relation to this model.

The experimental systems described here present a large amount of information at the search interface. There were concerns that this information would hinder subjects and lead to cognitive overload. In cognitive overload situations, a searcher's finite cognitive resources are stretched ever thinner by increased demands placed on them to process information. When faced with a plentiful supply of information traditional presentation strategies, searchers typically have to make a series of decisions: Is this title relevant? Are these terms in the correct context? What comes after the ellipses? Where are these snippets in the document? Is the surrogate relevant? Shall I click this title? Every decision has an associated cost: time, effort, and stress (Kirsh, 2000). The top-ranking sentences restrict the decisions searchers make to those about the *relevance* of the information: Is this sentence relevant? Shall I click the sentence?

Subjects in all studies were asked to comment on the search process they performed on each of the systems, in particular they were asked how stressful or relaxing the search process had been. Cognitive overload scenarios can create "information anxiety" (Wurman, 1989), where the searcher becomes overwhelmed by information and trapped between their current state of knowledge and the amount of knowledge they require to solve the problem that initiated their seeking. Kuhlthau (1991), suggests that anxiety is an intrinsic part of the search process and will not totally disappear until the searcher has successfully completed their task. However, it is possible to minimize this anxiety by providing levels of support that help searchers reach their goal. In the three studies, the presentation of more content at the results interface did not lead to high levels of stress reported by subjects during their search; generally subjects found the experimental systems intuitive. This is a worthwhile finding, as the benefits of top-ranking sentences could be nullified if searchers felt stressed using the systems.

Kuhlthau's model of the ISP is divided into six stages that describe the search from beginning to end: initiation, selection, exploration, formulation, collection, and presentation. Each stage has common affective, cognitive, and physical activities and require different levels of support from a search system. The systems described in this article support three of the six stages: exploration, formulation, and collection. The other stages are typically carried out before the

search system is used (understanding their information need and selecting search topics) or after the conclusion of the search (reporting the findings).

During the exploration stage searchers try to find information that will increase their understanding of what information is needed to complete their search. Kuhlthau (1991) suggests that during the exploration stage, strategies ". . . which open opportunities for forming new constructs such as listing facts which seem particularly pertinent . . . may be helpful during this time" (p. 366). The top-ranking sentences are a list of query-relevant document representations that may help searchers better understand their information need and begin conceptualizing these needs to form search statements.

The systems presented here provide limited support for the formulation stage of the ISP. This assumes that there is a point of "focus" (Belkin, 1980; Kelly, 1963; Kuhlthau, 1991) where uncertainty drops and searchers can better identify the topic of their search. During this stage, searchers formulate a focus during which they better understand their information need and the information they are searching for. The formulation stage is personalized and search systems that fully support it help searchers construct query statements. In the systems described in this article, it is the system's internal representation of the information need that changes when presented with relevance information. This is hidden from the searcher, who only sees the effect of the revised formulation, i.e., the reordered list of top-ranking sentences. The systems support the improvement of search queries but since there is no direct dialogue with the searcher about these new queries their support for the formulation stage of the ISP is limited.

The experimental systems may also be useful during the collection stage of the ISP. The presentation of top-ranking sentences gives searchers an opportunity to examine search results more closely and gather pertinent information from a variety of information sources. The search statements created as "focus" was obtained are improved and enhanced (internally) and used to reorder the top-ranking sentence lists during the search.

Top-Ranking Sentences

The top-ranking sentences were generally well received by experimental subjects. Although, from the user studies it did emerge that the training task and orientation sessions were important as searchers initially expressed concerns about the unfamiliarity of the interface. In this section we discuss subject perceptions of the TRS under three main section headings: Task, Popularity, and Usability.

Task. There were variations in the performance of top-ranking sentence-based interfaces for different types of search task in the TRSPresentation and TRSDocument studies. Subjects felt that background and decision tasks required information from a number of sources to get a general overview of a topic or to make reasonable search decisions. The top-ranking sentences were effective at facilitating access to such information. However, for the fact searches the top-ranking sentences were not perceived as being as useful. That is, when searchers were fully aware of what they were looking for, they felt that they did not require additional interface support, and that they would be best able to find useful information with the commercial search engine they used most frequently. This does not imply that the top-ranking sentences were useless; they were simply not required for the completion of this type of search task.

Popularity. Any problems experienced by subjects were mainly related to their unfamiliarity with top-ranking sentence-based interfaces. To interact well with the systems presented in these studies, subjects had to change the way they searched for useful information. The approach encouraged more examination of search results and a reduction in the number of query reformulations; a shift from the well-established search paradigm currently promoted by Web search engines. The negative findings above do not express a dislike for top-ranking sentences, but for *any* change in the way results are presented. This may also suggest that if subjects are confident about being able to find information before starting to search they would rather use a familiar system (i.e., one where they do not have to think much about the interaction or the interface itself).

The value of titles, sentence fragments, and URLs used by traditional Web search engines were tested in TRSPresentation. Searchers use these surrogates to make decisions about which documents to download and view. The user studies demonstrated that subjects rarely use interface features such as the "Next" button (all studies) or the URL of the document (TRSPresentation).[4] In the top-ranking sentence systems the URL and the Next button, although present, were not regarded as being as important.

Across all studies, the sentences and associated interface features were liked by subjects. In TRSPresentation we shifted the focus from document surrogates to the actual content of the document. In doing this, we found that the document titles were less useful as searcher attention was drawn to the information resident inside documents. The experimental system used in TRSPresentation increased awareness of returned document set content, allowing searchers to make better decisions on the relevance of both the retrieved set of documents and documents individually.

Usability. In the experimental systems that presented results as a ranked list of documents, subjects would rather reformulate and resubmit their queries than deeply peruse the documents returned to them. In doing so, they may discard potentially relevant documents without giving them due consideration. The document list returned is only an algorithmic match to the searcher's query, something that typically contains only one or two query terms (Jansen et al., 2000). Unless

---

[4]This was the only study where we measured the usefulness of the URL.

TABLE 3. Percentage difference between top-ranking sentence systems and experimental (ranked document) baselines.

| | Experimental factor | | | | |
| | Page views | | | Task completion | |
Study	Overall	Outside first 10	Queries	Time	Number of tasks
TRSDocument	+43.59	+76.46	−38.80	−8.50	+16.67
TRSPresentation	+65.41	+115.44	−61.20	−8.68	+18.32

the information need is very specific (i.e., someone's name, such as in the fact search) the system may struggle to provide a ranking that is a match for the searcher's information need. This problem is amplified if the system only ranks whole documents as small highly relevant sections may reside in documents with a low overall ranking.

The top-ranking sentences encourage more interaction with the retrieved document set, lowered the number of queries submitted, and improved task success. Table 3 shows the percentage differences with the experimental baselines used in the TRSPresentation and TRSDocument studies. If more than one top-ranking sentence system is used in the study or there is more than one non-TRS baseline then results are averaged across systems.

As can be seen from Table 3, the top-ranking sentences encourage more page views outside the top 10 documents, more page views in general and a reduced number of query iterations. The increased number of page views coincided with a greater sense of task completion. The reduced number of queries suggests that searchers were interacting in a way symptomatic of increased perusal with the returned set. The shorter task completion times and increased number of tasks completed suggests that searchers were using their time more efficiently.

Implicit Feedback

The traditional view of information seeking assumes a searcher's need is static and represented by a single query submitted at the start of the search session. However, as is suggested by Harter (1992) among others, the need is in fact, dynamic and changes to reflect the information viewed during a search. As they view this content their knowledge changes and so does their problematic situation. It is therefore preferable to express this modified problem with a revised query. The experimental systems in TRSFeedback and TRSDocument do this, selecting the most useful query modification terms during a search.

In the systems developed in these studies, the sentences are reordered using implicit relevance information gathered unobtrusively from searcher interaction. Experimental subjects found this a useful feature that helped them find relevant information. They suggested that it was most useful when they felt the initial query had retrieved a large amount of potentially relevant information and they wanted to focus

their attention on only the most relevant parts. These systems are more push-oriented than the static top-ranking sentences system tested in TRSPresentation. The systems are adaptive, work to better represent information needs and consider changes in these needs, restructuring the content presented at the results interface.

In TRSFeedback and TRSDocument we assumed that the viewing of a document's summary was an indication of an interest in the relevance of the summary's contents. There are several grounds on which this can be criticized: Searchers will view nonrelevant summaries; the title rather than the summary was what the user expressed an interest in; and the searcher may look at all retrieved documents before making real relevance decisions. Nevertheless, we felt that this assumption was fair enough to allow an initial investigation into the use of implicit feedback. In TRSDocument we introduced a timing mechanism to eliminate the problems caused by the accidental 'mouseover' of document titles and the unwanted removal of sentences from the top-ranking sentences list that follows. The results of TRSDocument are testament to the success of a very limited version of an implicit feedback technique.

Despite positive feedback, subjects also had two reservations regarding how implicit feedback was used in the system. First, as the reordering occurred at the same time as a summary appeared or updated, they did not always notice the effect of the reordering. The presentation of the updating therefore needs improving in future systems. Second, the top-ranking sentences only contained sentences from Web pages for which the searcher had not already viewed a summary. If the searcher viewed the summary for a page, then all sentences from that page were removed from the list of top-ranking sentences. This choice was made to increase the degree to which the list of top-ranking sentences would update. However, many subjects stated that they would prefer less updating and no removal of sentences.

The results of the three studies show that it is possible to get searchers to interact with more than a few search results. The approach moves away from simply presenting titles to presenting alternative access methods for assessing and targeting potentially relevant information. The findings are useful for the development of search interfaces to help users of search systems search more effectively. More complex and effective techniques based on these findings are described in subsequent work (White, Jose, & Ruthven, 2004).

Conclusions

In this article we have introduced a content-driven information seeking approach and described three studies to test its effectiveness. The studies are related, each adding additional interface support, but all using top-ranking sentences. In the first, we used the ranked sentences as an alternative to document lists, shifting searcher attention from the document surrogates to the document content. The second used the sentences to reflect the use of two contrasting relevance feedback techniques. The third used the sentences to encourage

interaction with the retrieved set, to reflect the dynamic nature of information needs and to complement, rather than replace, document lists. Each study involved real searchers and different types of information seeking scenario.

We have introduced and described push and pull information seeking and explained how these approaches differ. Top-ranking sentences are a precision-oriented approach that increase the amount of useful information a searcher can access. We have shown that this approach, whether or not supported by additional implicit feedback techniques, can lead to effective and efficient searching.

Ranking documents is a heavy-handed, cumbersome means of result presentation. Documents may not be entirely relevant and document surrogates may not be strictly indicative. It is the information in the documents that searchers seek. Our approach extracts, ranks, and presents the content of the returned set, blurring interdocument boundaries and encouraging information seeking based on the pertinent document content.

Acknowledgments

References

Ahlberg, C., & Shneiderman, B. (1994). Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In B. Adelson, S. Dumais, & J. Olson (Eds.), Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (pp. 313–317). New York: ACM.

Amitay, E., & Paris, C. (2000). Automatically summarising web sites: Is there a way around it? In A. Agah, J. Callan, E. Rundensteiner, & S. Gauch (Eds.), Proceedings of the 9th International Conference on Information and Knowledge Management (pp. 173–179). New York: ACM.

Belkin, N.J. (1980). Anomalous state of knowledge for information retrieval. Canadian Journal of Information Science, 5, 133–143.

Belkin, N.J. (1984). Cognitive models and information transfer. Social Science Information Studies, 4, 111–129.

Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982). Ask for information retrieval: Part i. Background and theory. Journal of Documentation, 38(2), 61–71.

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

Chen, H., & Dumais, S.T. (2000). Bringing order to the web: Automatically categorizing search results. In T. Turner & G. Szwillus (Eds.), Proceedings of the ACM SIGCHI Conference on Human Factors in Computing System (pp. 145–52). New York: ACM.

Claypool, M., Le, P., Waseda, M., & Brown, D. (2000). Implicit interest indicators. In D. Riecken, D. Benyon, & H. Lieberman (Eds.), Proceedings of the 6th International Conference on Intelligent User Interfaces (pp. 33–40). New York: ACM.

Cool, C., Park, S., Belkin, N.J., Koenemann, J., Ng, K.B. (1996). Information seeking behaviour in new searching environment. In P. Ingwersen & N.O. Pors (Eds.), Proceedings of Second International Conference on Conceptions of Library and Information Science (pp. 403–416). Copenhagen: Royal School of Librarianship.

Crouch, C.J., Crouch, D.B., Chen, Q., & Holtz, S.J. (2002). Improving the retrieval effectiveness of very short queries. Information Processing and Management, 38(1), 1–36.

Dumais, S.T., Cutrell, E., & Chen, H. (2001). Optimizing search by showing results in context. In J. Jacko & A. Sears (Eds.), Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (pp. 277–284). New York: ACM.

Dziadosz, S., & Chandrasekar, R. (2002). Do thumbnail previews help users make better relevance decisions about web search results? In K. Jarvelin, M. Beaulieu, R. Baeza-Yates, & S.H. Myaeng (Eds.), Proceedings of 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 355–356). New York: ACM.

Edmundson, H.P. (1964). Problems in automatic abstracting. Communications of the ACM, 7(4), 259–285.

Furnas, G.W. (1997). Effective view navigation. In S. Pemberton (Ed.), Proceedings of the ACM SIGCHI Conference on Human Factors in Computer Systems (pp. 367–374). New York: ACM.

Furner, J. (2002). On recommending. Journal of the American Society of Information Science and Technology, 53(9), 747–763.

Goecks, J., & Shavlik, J. (2000). Learning users' interests by unobtrusively monitoring their normal behavior. In D. Riecken, D. Benyon, & H. Leiberman (Eds.), Proceedings of the 2000 International Conference on Intelligent User Interfaces (pp. 129–132). New York: ACM.

Harter, S.P. (1992). Psychological relevance for information science. Journal of the American Society for Information Science, 43, 602–615.

Hearst, M. (1995). Tilebars: Visualization of term distribution information in full text information access. In R. Katz, T. Mack, L. Marks, M.B. Rosson, & J. Neilson (Eds.), Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (pp. 59–66). New York: ACM.

Hemmje, M. (1995). Lyberworld: A 3d graphical user interface for fulltext retrieval. In J. Miller, I. Katz, R. Mack, & L. Marks (Eds.), Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (Conference companion, pp. 417–418). New York: ACM.

Hersh, W., & Over, P. (2001). Trec-10 interactive track report. In E.M. Voorhees & L.P. Buckland (Eds.), Proceedings of the Tenth Text Retrieval Conference (pp. 38–41). Gaithersburg, MD: Department of Commerce, National Institute of Standards and Technology.

Ingwersen, P. (1994). Polyrepresentation of information needs and semantic entities: Elements of a cognitive theory for information retrieval interaction. In B. Croft & C.J. Van Rijsbergen (Eds.), Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 101–110). New York: ACM.

Jansen, B.J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the web. Information Processing & Management, 36(2), 207–227.

Kelly, G.A. (1963) A theory of personality: The psychology of social constructs. New York: Norton.

Kirsh, D. (2000). A few thoughts on cognitive overload. Intellectia, 30, 19–51.

Kuhlthau, C. (1991). Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science, 42(5), 361–371.

Paice, C.D. (1990). Constructing literature abstracts by computer: Techniques and prospects. Information Processing & Management, 26(1), 171–186.

Robertson, S.E. (1990). On term selection for query expansion. Journal of Documentation, 46(4), 359–364.

Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 41(4), 288–297.

Saracevic, T. (1975). Relevance: A review of and a framework for thinking on the notion of information science. Journal of the American Society for Information Science, 26(6), 321–343.

Spink, A., Griesdorf, H., & Bateman, J. (1998). From highly relevant to not relevant: Examining different regions of relevance. Information Processing & Management, 34(5), 599–621.

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

Tombros, A., & Sanderson, M. (1998). Advantages of query-biased summarisation in information retrieval. In W.B. Croft, A. Moffat, & C.J. Van Rijsbergen (Eds.), Proceedings of the 21st Annual ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 2–10). New York: ACM.

White, R.W., Jose, J.M., & Ruthven, I. (2003a). A granular approach to web search result presentation. In G.W.A. Rauterberg, M. Menozzi, & J. Wesson (Eds.), Proceedings of the 9th IFIP TC13 Conference on Human Computer Interaction (pp. 213–220). Amsterdam: IOS Press.

White, R.W., Jose, J.M., & Ruthven, I. (2003b). A task-oriented study on the influencing effects of query-biased summarisation in web searching. Information Processing & Management, 39(5), 707–733.

White, R.W., Jose, J.M., & Ruthven, I. (2004). An implicit feedback approach for interactive information retrieval. Information Processing & Management, in press.

White, R.W., Ruthven, I., & Jose, J.M. (2002a). Finding relevant web documents using top ranking sentences: An evaluation of two alternative schemes. In K. Järvelin, M. Beaulieu, R. Baeza-Yates, & S.H. Myaeng (Eds.), Proceedings of the 25th Annual ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 57–64). New York: ACM.

White, R.W., Ruthven, I., & Jose, J.M. (2002b). The use of implicit evidence for relevance feedback in web retrieval. In F. Crestani, M. Girolami, & C.J. Van Rijsbergen (Eds.), Proceedings of 24th BCS-IRSG European Colloquium on Information Retrieval Research (pp. 93–109). Berlin: Springer-Verlag.

Wurman, R.S. (1989). Information anxiety. New York: Doubleday.