Part II

Facilitating Effective Information Access

So far in this thesis I have introduced information retrieval (IR), relevance feedback (RF) and implicit feedback measures for IIR. In this part an approach is proposed to facilitate searcher interaction with the retrieved documents through the use of document representations such as query-relevant *Top-Ranking Sentences* extracted from Web documents. I call this approach 'content-driven information seeking' and it tries to encourage more interaction with search results. The approach is evaluated in three related user studies, and the findings discussed. Motivated by the success of these techniques in the user studies, I also extend this work and present an overview of a search interface that uses these techniques to present these representations to searchers and allows them to follow interactive *relevance paths* between them.

Chapter 3

Top-Ranking Sentences

3.1 Introduction

Query-relevant *Top-Ranking Sentences* chosen from top-ranked retrieved search results are used as an interface component to assist searchers throughout this thesis. These sentences are selected based on the searcher's query, facilitate access to potentially relevant information and encourage a deeper examination of search results. Documents returned in response to a query by the search system are used to create the Top-Ranking Sentences. [5] These documents are downloaded and all sentences from each document are extracted. Each sentence is assigned a score, using the scoring methodology described later in this chapter. This uses factors such as position of the sentence in document, the presence of any emphasised words and any terms that occur both in the sentence and the document title. In addition sentences receive additional scores depending on the proportion of query terms they contain. This component ensures the scoring mechanism treats sentences that use query words as important.

In this chapter I describe the Top-Ranking Sentences, give the reasons why sentences, and not other semantic entities, such as paragraphs, were chosen, and provide details on how sentences were extracted and scored. It is possible to use different presentation strategies to show these sentences to the searcher; this chapter begins with a description of the strategies used.

3.2 Presentation Strategies

Two presentation strategies are adopted in the interfaces described in this thesis: sentences combined to form a summary for each document and as a list across documents.

[5] The sentences selected are therefore dependent on the document ranking algorithms used by the underlying search system.

3.2.1 Sentences as Document Summary

The Top-Ranking Sentences are chosen from each document and are presented at the interface for each document. The sentences combine to form a summary of the document. In response to a searcher's query, Web search engines typically only present results that consist of *document surrogate* information such as short sentence fragments and meta information similar to that shown in Figure 3.1.


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

Figure 3.1. Web search engine result for the query 'dust allergies'.

Search engines such as Google use query-biased techniques e.g., (Tombros and Sanderson, 1998) to select these sentence fragments and present query terms in the context they occur in the document. To provide this context, such systems use leading and trailing non-query terms to create short snippets of text centred on the query. These snippets, separated by ellipses, are combined to construct the document summary. This information – along with document title and the uniform resource locator (URL) – is used by searchers when deciding which documents to visit. The importance of showing searchers clues of the information resident in the source document has already been established in Landow's work on *rhetoric of departure* (Landow, 1987) and Furnas's work on *information scent* or *residue* (Furnas, 1997). Figure 3.2 shows one way in which these Top-Ranking Sentences can be used to form a summary of a retrieved document.

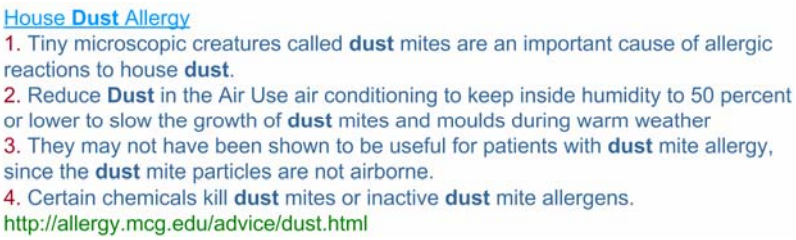
House **Dust** Allergy
1. Tiny microscopic creatures called **dust** mites are an important cause of allergic reactions to house **dust**.
2. Reduce **Dust** in the Air Use air conditioning to keep inside humidity to 50 percent or lower to slow the growth of **dust** mites and moulds during warm weather
3. They may not have been shown to be useful for patients with **dust** mite allergy, since the **dust** mite particles are not airborne.
4. Certain chemicals kill **dust** mites or inactive **dust** mite allergens.
http://allergy.mcg.edu/advice/dust.html

Figure 3.2. Sentences as document summary for the query 'dust allergies'.

In an earlier user study I demonstrated that using the best four Top-Ranking Sentences as a Web document summary was preferred to the presentation strategies exemplified in Figure 3.1 (White *et al.*, 2003b). In this user study, I found that the increased information allowed searchers to make more reliable relevance assessments, experience more satisfying searches and search more effectively. This presentation strategy groups sentences based on their

source document. However, it is also possible to present Top-Ranking Sentences in a ranked list, independent of source document. In the next section I describe this approach.

3.2.2 Sentences as List

Presenting Top-Ranking Sentences independent of source documents allows highly relevant sentences from lower ranking documents, which may never be viewed simply because of their resident document's rank position, to be made accessible to the searcher. Figure 3.3 shows part of a list of Top-Ranking Sentences taken from one of the three user studies described in Chapter Four.

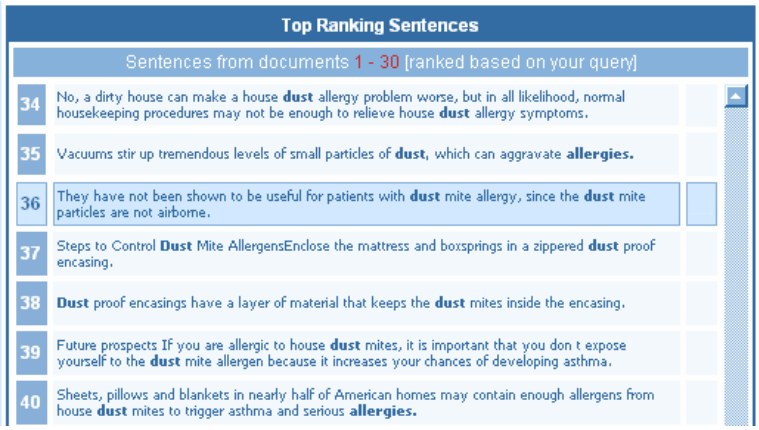

Figure 3.3. A portion of a list of Top-Ranking Sentences for the query 'dust allergies'.

The sentences are numbered based on their rank position and shown individually in the list, with query terms highlighted.

Presenting sentences in this way provides a high level of granularity, removing the restriction of document boundaries and shifting the focus from the document as a semantic entity to the information the document contains. This means that searchers are not forced to access information through documents but through the actual content of documents. Through ranking this information with respect to the query, the searcher is given an overview of the content of the returned set. A document list is biased towards the searcher's information need at the document level; documents that are a close match to the searcher's query appear near the top of the list. Presenting lists of Top-Ranking Sentences biases at the sentence level; sentences that are a close match to the searcher's query are shown near the top of a ranked list of sentences. As will be described in Chapter Four the sentences can also be used to facilitate

access to low-ranked documents and communicate the effects of relevance feedback decisions.

In this thesis both sentence presentation strategies are used to assist searchers. In the next section I explain why sentences were chosen as an interface component.

3.3 Why Sentences?

Earlier studies have shown that using semantically richer document representations can be beneficial to searchers and allow them to make more reliable relevance assessments (Spink *et al.*, 1998; White *et al.*, 2003b). In this thesis sentences are used as a component to construct representations of documents that encourage searchers to examine search results more closely.

The rationale behind sentence extraction is to find a subset of the source document that represents its contents or the query, typically by scoring words and then sentences according to specific rules. The rules mainly concern the identification of clues for the importance of each sentence in the source document. Sentence extraction methods are capable of producing acceptable summaries that are domain independent (Luhn, 1958; Edmundson, 1969; Rush *et al.*, 1971; Paice, 1981; Brandow *et al.*, 1995; Salton *et al.*, 1997). This makes them perhaps more suitable for heterogeneous collections such as the Web than language generation (McKeown *et al.*, 1995) or artificial intelligence (Tait, 1985) techniques that display only a marginal level of usefulness within their restricted domains.

Research on automatic sentence extraction is well-documented. In the approach described in this chapter, sentences were used as interface components for two reasons: (i) they are by definition a coherent linguistic entity to overcome problems with semantics and present the query terms in context, (ii) they are small enough to allow searchers to assess relevance in a short time. These are preferred to paragraphs (as used in passage retrieval (Salton *et al.*, 1993; Callan, 1994)) simply because they take less time to assess. This allows searchers to make speedy judgements on the relevance/irrelevance of the information presented to them. Sentences are also the preferred semantic entity for analysis and retrieval in linguistic-based IR (Smeaton, 1990) and in the Novelty Track at the Text Retrieval Conference (TREC) (Harman, 2002).

Sentences are also used in multi-document summarisation approaches, where sentences pooled from a number of documents are used to provide a summary of these documents. Such summaries are relatively short, use domain-specific methods to score sentences (Radev

and McKeown, 1998) and place a strong emphasis on coherence (Goldstein *et al.*, 2000). Sentences can also be used to form summaries of Web document clusters, as one application of the methods described in this chapter suggests (Osdin *et al.*, 2002).

In the next section I describe how sentences were selected by the search system.

3.4 Selecting Sentences

To form a list of Top-Ranking Sentences I use a sentence extraction model similar to that proposed by Tombros and Sanderson (1998). The approach extracts sentences from the top-ranked Web documents retrieved in response to a searcher's submitted query. The Web was used as searchers had experience interacting with Web documents, effective baseline search systems were readily available and realistic search scenarios for user evaluations could be easily created.

This section describes the sentence selection architecture and the techniques used to extract and score candidate sentences. Figure 3.4 shows a general overview of the approach used.

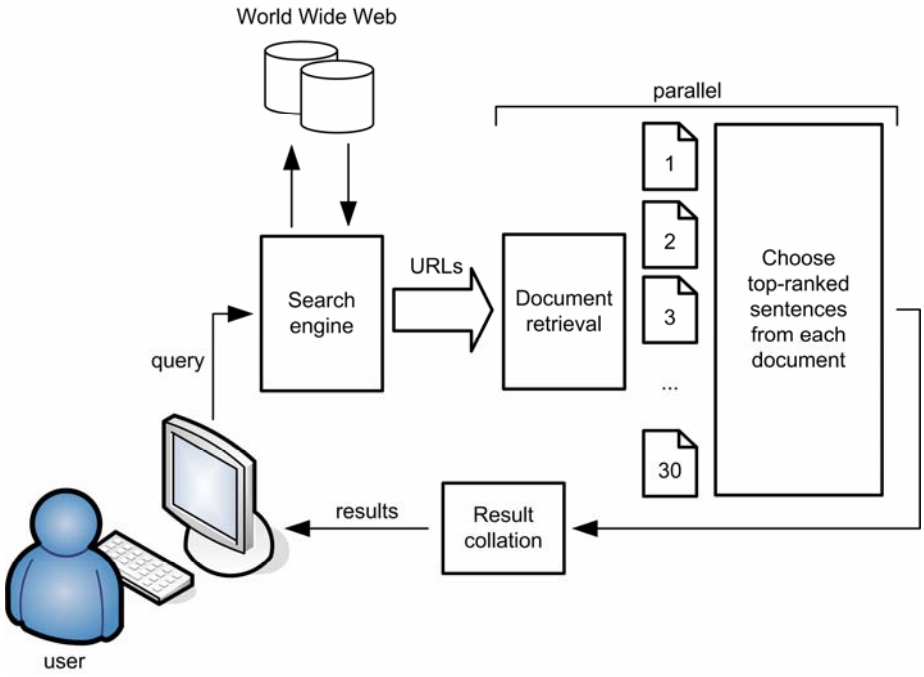

Figure 3.4. Top-Ranking sentence selection architecture.

A searcher's query statement is first passed to a Web search engine, which returns a set of documents. The documents are then visited by the system in parallel and the resident sentences extracted. The sentences are scored according to how useful they will be in

reflecting page content and relevance assessment. Sentence extraction has been shown to have useful applications in Web document summarisation (Berger and Mittal, 2000). Extraction mechanisms are useful for selecting the potentially useful parts of Web documents as they can handle small portions of information and are domain independent. The extraction methods used standard punctuation (e.g., full stop, exclamation mark and question mark) and first character capitalisation methods to determine where sentences start and stop.

3.4.1 Sentence Scoring

Sentences are scored based on four criteria; *title* (e.g., sentence terms that co-occur with the title), *location* (e.g., where a sentence resides in a document), *relation to query* (e.g., the proportion of query words a sentence contains) and *text formatting* (e.g., the additional formatting added by the document author). Each scoring method is now described.

3.4.1.1 The Title Method

This method assumes the author of a document reveals the main concepts in the title of their work. It also assumes that when an author divides his work into sections, he does so in a standard manner, selecting appropriate headings for each of these divisions. Sentences containing terms that appear in the title and headings are given more weight than those without. Edmundson (1969) experimented with this method using a collection of technical documents, and assigned a greater importance to terms that appear in the title than in the section headings. The final sentence score for each sentence could then be found through the sum of the weights of each title word in the sentence. It was thought reasonable to use this method to score the sentences in Web documents as the document author has control over the title of the document and the content of the page. The title may not provide enough information on its own or supplemented with other meta-information (as in traditional result lists) to be truly indicative, but it may contain some important keywords.

3.4.1.2 The Location Method

This method assumes that: (i) that sentences located under certain headings in a document convey significant content and are therefore relevant, (ii) that important sentences tend to occur near the start, or near to the end, of a document and its paragraphs (Edmundson, 1969; Brandow *et al.*, 1995). This method assigns positive weights to words occurring under headings in a document (represented by the `<H1>...<H6>` HTML [6] tags) and computes the *heading weight*. As well as this, the method also assigns weights to sentences based on their ordinal position in the document (the *ordinal weight*), i.e., the first and last paragraphs in the

[6] HyperText Markup Language (HTML).

document and the first and last sentences in the paragraphs. Paragraph termination is detected in Web documents using instances of the </P> and

 HTML tags. The total location method score for a sentence is the combination of the heading and ordinal weights.

3.4.1.3 The Text Formatting Method

The rationale behind this method stems from the idea that a Web document author may emphasise important terms (or keywords) in some way. When using the HTML that most Web documents are written in, the author can format text in a number of ways, such as **bolded**, *italicised* and underlined.

When formatted terms occur in a sentence, the sentence score is incremented by a small amount for each term. The values used were chosen based on beliefs about the value of this evidence and through pilot testing. If a term is formatted in two or more ways, say bold and italic, then the score for that sentence is incremented for each piece of formatting separately.

3.4.1.4 The Query-Biased Method

This method assumes that if searchers could see the sentences in which their query terms appeared they would be able to make a better assessment of document relevance. Tombros and Sanderson (1998) proposed a method for calculating a query score for each sentence in the document, based upon its relevance to the query. The larger the number of query terms in a sentence, the more relevant the sentence is likely to be.

The top scoring sentences are selected until the desired number of sentences is reached. This is defined to be 15-20% of the document length, or a maximum of four sentences and concurs with previous work (Edmundson, 1964; Brandow *et al.*, 1995; Kupiec *et al.*, 1995).

A potential drawback of using query-biased approaches to summarise *documents* is the biased view of the document that results; only those sentences containing many query terms are promoted. The resultant effect is a representation of the document that may not be indicative of the actual document and the emphasis therein. This problem is made more acute if the documents contain information on a variety of topics, one of which happens to be the topic of the need. Paice (1990) refers to this as the 'coverage and balance' problem, and is a flaw of the extracting approach. Also, it is possible that sentences containing the query terms can be scattered throughout the document. Document summaries composed of these sentences may have no cohesion and simply represent as much of the text as possible (Amitay and Paris, 2000).

3.4.1.5 Summary of Methods So Far

So far in this chapter I have described four heuristic-based methods to score the sentences extracted from Web documents. I conducted a pilot test to evaluate the sentences chosen by this approach and combined the best Top-Ranking Sentences from each document to form a document summary. Joining the sentences in this way is only one possible use of Top-Ranking Sentences and other applications are described in later chapters of this thesis. Summaries were presented to subjects as part of an interface to the Google [7] and AltaVista [8] search systems and compared with traditional forms of result presentation, where lists of titles, sentence fragments and URLs (similar to Figure 3.1) were presented. Subjects found the enriched summaries useful and that it encouraged them to interact with their search results more closely (White *et al.*, 2003b). However, the pilot study also revealed some minor problems, namely:

i. *Some sentences were too short.* Some highly scoring sentences were often headings that had been incorrectly labelled by the document author (i.e., not inside the appropriate tags). These sentences were too short to be indicative.
ii. *Some sentences were redundant.* The four Top-Ranking Sentences from each document were often too similar, query terms were shown in similar contexts and the value of the summary generated was diminished.

As a result, I incorporated two more methods to improve the quality of the sentences selected. These are *sentence length cut-off* and *redundancy checking*.

3.4.1.6 Sentence Length Cut-off

This method addressed problems with selecting sentences that were too short. All sentences used by the scoring methods need to be of a certain length (threshold: 15 tokens including punctuation). This is a frequently used threshold for removing captions, titles and headings (Kupiec *et al.*, 1995; Teufel and Moens, 1997). These headings are handled separately in the approach described in this chapter (see Section 3.4.1.2).

3.4.1.7 Redundancy Checking

To address problems with sentence redundancy a means of redundancy checking was used when selecting Top-Ranking Sentences. Through combining query-biased methods and techniques for reducing the level of redundancy it may be possible to select sentences that are

[7] http://www.google.com
[8] http://www.av.com

query-relevant and show the query terms in different contexts, one of which may be useful for the searcher. This can help ensure that sentences are selected in relation to the query that can also provide an overview of retrieved information.

The redundancy checking techniques used are based on those of Gong and Lui (2001). Unlike their work I do not use term frequency vectors for each document and compute the similarity to the document's vector. Since the approach does not create a generic document summary, there is no need to compute the similarity to the document. However, the approach does compute the degree of similarity to the query. The technique used is illustrated in Figure 3.5.

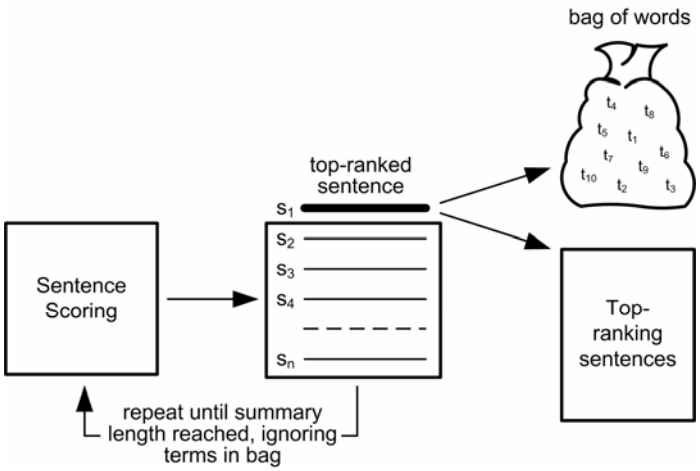

Figure 3.5. Redundancy checking in sentence selection.

The sentences extracted from the Web documents are scored based on the initial searcher query and all other methods described so far in this chapter. The sentences are then ranked based on these scores and the top sentence is removed and stored as a 'top-ranking sentence'. The *non-query* words from this sentence are placed in a bag and the process repeats, i.e., all sentences (except the one that was removed) are rescored and reordered using all constituent words that are not in the bag. The sentences chosen by this method are those that represent the query terms in different document contexts. This makes the sentences chosen suitable for document content overview (when grouped per document) or result set overview (when grouped across all top-ranked documents).

3.4.2 Combining Sentence Scores

The methods above are applied to a sentence in the sequence shown in Figure 3.6. This results in a final sentence score. The final sentence score is computed by summing together

all scores from all methods. The inclusion of this scoring method had no detrimental effect on the overall sentence score should a title word *not* occur in a sentence, but a benefit if it does. All methods are given an opportunity to weight sentences; in reality a large proportion of a sentence's score is derived from its relation to the query. The redundancy checking uses all sentence scoring methods but operates independently of them and is therefore not included in the figure. The sentence length cut-off acts as a filter prior to any scoring to aid system efficiency, since only sentences of sufficient length will eventually be scored.

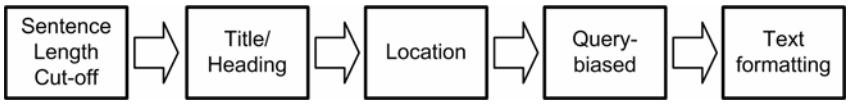

Figure 3.6. Sentence scoring methodology.

A drawback of applying a linear combination of the methods identified above is the implication that the clues provide independent items of evidence that simply needs to be combined. This may not be true, as it may be possible for the clues to interact in some way. For example, a term that is bold, underlined and in the title of the document should perhaps contribute more to its residing sentence's score than the sum of the scores for the title-keyword and twice for the text formatting (bold and underline). Despite this drawback, many studies (Edmundson, 1969; Kupiec *et al.*, 1995; Tombros and Sanderson, 1998; White *et al.*, 2003b) have used this cumulative technique to good effect for selecting sentences. In the approach presented in this thesis the chosen sentences can be used to create summaries of documents and other document representations, and presented in a ranked list, independent of source document.

3.4.3 Error Handling

The top-ranking sentence selection architecture illustrated in Figure 3.4 may experience problems selecting sentences from Web documents. This could be for a number of reasons; the document contains HTML frames, contains little or no text, or takes too long to download. [9] If this happens, or if a document is one of the restricted document types [10] then the top-ranking sentence selection architecture tries to choose sentences from the search engine's cached version of the page. The strategy employed if this is unsuccessful is dependent on the presentation strategy. In the 'Sentences as document summary' approach,

[9] The top-ranking sentence selection system rejects a Web document if it takes more than 3 seconds to download.

[10] For technical reasons, the techniques cannot select Top-Ranking Sentences from proprietary non-text files e.g., Microsoft Word documents (.doc), Microsoft Excel spreadsheets (.xls), PostScript files (.ps) and Adobe Portable Document Format files (.pdf).

the small collection of sentence fragments taken from the search engine (such as that shown in Figure 3.1) is used as a pre-created alternative to that created by the system. In contrast, in the 'Sentences as list' approach, the sentence fragments from the search engine are treated as a single sentence and included in the list of Top-Ranking Sentences as an additional entry.

3.4.4 Other Sentence Selection Methods

It is worth noting that other methods exist for selecting sentences extracted from documents. The *keyword method* (Luhn, 1958) assumes that high-frequency words that are not common *stop words* (e.g., 'of', 'the', 'and') are indicative of the document's content and are therefore useful for scoring sentences. Rather than assigning a weight to each term according to the number of times it occurs, as in (Rath, 1961; Earl, 1970), the method involves locating clusters of significant words within sentences and assigning scores to them accordingly. The query-biased approach is a version of the keyword method. Instead of providing a list of candidate index terms for each document that refer to the central concepts of the document, the searcher provides the retrieval system with a list that reflects the central concepts of the information need as they perceive it. This way, the sentences obtained from each document are those with a high score in relation to the searcher's expressed information need and have a high likelihood of relevance. The use of *syntactic criteria* (Earl, 1970), the *cue method* (Edmundson, 1969; Rush *et al.*, 1971) and the *indicator-phrase method* (Paice, 1990) rely on detailed knowledge of the corpus's language constructs and are therefore not appropriate for the heterogeneity of the Web. Paice (1990) and Spärck-Jones and Endres-Niggermeyer (1995) provide a thorough review of previous work in automatic sentence selection.

3.5 Chapter Summary

In this chapter I have introduced Top-Ranking Sentences as an interface component to present search results and encourage access to retrieved information. The rationale behind using sentences has been given, as have the techniques used to score sentences. Top-Ranking sentences can be used as document summaries, to provide an overview of the result set content and assist searchers in locating useful information. In Chapter Four I describe three user studies that use these sentences as a replacement for document lists, to communicate the effects of relevance feedback decisions and to facilitate access with retrieved documents.

Chapter 4

Content-Driven Information Seeking

4.1 Introduction

In this chapter I describe an approach that uses the techniques introduced in the previous chapter to encourage a deeper examination of the contents of the document set retrieved in response to a query. The approach shifts the focus of perusal and interaction 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. Traditional search interfaces assume searchers examine results document-by-document. In contrast the approach proposed extracts, ranks and presents the *contents* of the top-ranked document *set*. *Top-Ranking Sentences* (TRS) extracted from top documents at retrieval time are used as fine-grained representations of document content and, when combined in a ranked list, an overview of these documents. In some of the systems described in this chapter, the interaction of the searcher provides implicit relevance feedback that is used to reorder the sentences where appropriate. This chapter serves as an introduction to the use of implicit feedback in this thesis and to the style of interfaces I create.

Three related user studies with 58 different subjects were carried out to test the effectiveness of using TRS to assist searchers and communicate relevance feedback decisions. The findings of these studies were important since they influence the design of systems described in later chapters. In the analysis of the findings I focus on the relationship between the studies and qualitative subject perceptions of the approaches I describe. Hereafter I refer to the three studies as *TRSPresentation*, *TRSFeedback* and *TRSDocument*. [11] 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 not

[11] *TRSPresentation* (Top-Ranking Sentences for result presentation), *TRSFeedback* (Top-Ranking Sentences for feedback decisions) and *TRSDocument* (Top-Ranking Sentences for document access).

presented in this chapter, only the subject perceptions of the techniques employed. The quantitative findings for all three studies can be found in White *et al.* (2003a) (*TRSPresentation*), White *et al.* (2002b) (*TRSFeedback*) and White *et al.* (2002a) (*TRSDocument*). This chapter describes how subjects use top-ranking sentence interfaces for their search, how this differs from traditional search methods and reason why top-ranking sentence interfaces are preferred over traditional forms of result presentation. The findings of these studies motivate the research presented in the remainder of this thesis. In the next section I describe two contrasting information seeking strategies for interacting with search interfaces; one encouraged by traditional search systems and another by systems that implement aspects of the content-driven paradigm I propose.

4.2 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 behaviour 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 (1-2 line) sentence fragments. These fragments normally contain at least one instance of the query terms and give the searcher an idea of the context in which the query terms are used in the document.

In result lists searchers assess document relevance externally, based on what they can infer from their surrogates. On the Web, authors assign document titles and the extent to which these titles are indicative of content can vary. This differs from the static homogeneous collections used in initiatives such as TREC (Voorhees and Harman, 2000), where there is consistency in the titles/headlines assigned. Figure 3.1 (in Chapter Three) showed an example of surrogate information used in search engine result lists. This information is important since searchers use it to make decisions about what documents to view (Furnas, 1997). 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 presenting full

sentences to the searcher, IR systems can present the query terms in the local context in which they are used within retrieved pages.

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 (Jansen *et al.*, 2000; Crouch *et al.*, 2002). The approach described in this chapter encourages more interaction with search interfaces and in some cases uses this interaction to make decisions on the searcher's behalf. I call this approach *content-driven information seeking* (CDIS) and it is in contrast to searcher-driven approaches where there is more onus on searchers to proactively seek information. In this section I introduce the concepts of *pull* and *push* information seeking; the latter encourages CDIS whereas the former does not.

4.2.1 Pull and Push Information Seeking

In this section two contrasting information seeking strategies are described: pull and push. The pull approach presents the searcher with surrogate document representations (e.g., titles, sentence fragments and URLs) and relies thereafter on the searcher to visit the document. In contrast, the push approach presents, and dynamically restructures, relevant content at the results interface, irrespective of source document. These strategies are affected by result presentation techniques that 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 I describe these information seeking strategies, and the differences between them.

4.2.1.1 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; this requires a document-by-document examination of search results. 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 judgements 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 the search proceeds. They are burdened with the responsibility to select additional query words and drive their own search. As suggested in Chapter Two, this can be problematic if the information need is vague (Spink *et al.*, 1998) or searchers are unfamiliar with the collection being searched or the retrieval environment (Salton and Buckley, 1990). The pull strategy is adopted by traditional search systems that, after the initial retrieval, require searchers to locate relevant information. In the next section I describe the contrasting push information seeking strategy.

4.2.1.2 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 inferred searcher interests. Two methods are used as enabling techniques for the push paradigm; Top-Ranking Sentences and implicit feedback. In this section I describe each of these.

4.2.1.2.1 Top-Ranking Sentences

Searchers can use the Top-Ranking Sentences, selected as described in the previous chapter, to guide them through their search. The Top-Ranking Sentences provide 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, meaning less time need be spent locating useful information, and more time can be spent assessing its value. These sentences can also be reordered using evidence gathered via implicit feedback from the searcher; in the next section I describe this process.

4.2.1.2.2 Implicit Feedback

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 and do not intrude on the searcher's primary line of activity i.e.,

satisfying their information needs (Furnas, 2002). In traditional relevance feedback systems, the function of making judgements is intentional, and specifically for the purpose of helping the system build up a richer body of evidence on what information is relevant. However, the ultimate goal of information seeking is to satisfy an information need, not to rate documents. Systems that use implicit feedback to model information needs and enhance search queries fit better with this goal.

As already mentioned in Chapter Two, implicit feedback systems typically use measures such as document reading time, scrolling and interaction to make decisions on what information is relevant (Claypool *et al.*, 2001). However, these systems typically assume that searchers will view and interact with relevant documents more than non-relevant documents. These assumptions are context-dependent and vary greatly between searchers. The approach used for implicit feedback in this chapter makes a potentially more robust assumption: searchers will try to view relevant information. Through monitoring the information searchers interact with search systems can approximate their interests. This is made possible since 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 relevance.

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 estimation of the searcher's interests. Since 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 in this chapter present subjects with search interfaces that may be unfamiliar to them. During these studies I 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. In the next section I compare the push and pull information seeking strategies.

4.2.2 Comparison of Information Seeking Strategies

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 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 4.1 shows other differences.

Table 4.1
Differences between the push and pull information seeking approaches.

Factor	Approach	
	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 relevance	Internally	Externally
System formulation of information needs	Static/Dynamic	Static

As Table 4.1 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 I 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. The findings of the studies influence the design of search interfaces described later in this thesis.

4.3 User Studies

Three user studies tested the worth of Top-Ranking Sentences in different information seeking contexts. The results from these studies are summarised in this chapter, each of which utilises 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 searcher interests 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. In each study Top-Ranking Sentences were taken from the top 30 documents to ensure the systems responded in a timely manner. In this section the generic experimental methodology is described, as are the experimental interfaces used, the tasks assigned and the relationship between studies.

4.3.1 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:

i. introductory orientation;
ii. pre-search/demographic questionnaire;
iii. 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 subject to attempt the task;

d. a post-search questionnaire;

iv. a final questionnaire, and;

v. an informal discussion (optional). [12]

There were minor differences in the methodology employed between studies, necessitated by the different experimental hypotheses.

4.3.2 Subjects

The recruitment of experimental subjects in these studies was specifically aimed at targeting two groups of subjects; *inexperienced* and *experienced*. Two out 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 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 systems.

4.3.3 Tasks

In *TRSPresentation* and *TRSDocument* subjects attempted combinations of tasks from the following categories: *fact* search (e.g., finding a named person's current email address), *decision* search (e.g., choosing the *best* impressionist art museum) and *background* search (e.g., finding information on dust allergies) (White *et al.*, 2002a). The tasks used are included in Appendix E. Each search task was placed within a simulated work task situation, (Borlund, 2000b) that created realistic search scenarios and allowed personal assessments of what information was relevant. *TRSFeedback* was carried out as part of the TREC 2001 Interactive Track (Hersh and 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.

[12] 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.

4.3.4 Interfaces

Each of the three studies used Top-Ranking Sentences to facilitate information access, encourage interaction and convey system decisions. In this section I describe the interfaces used in each study and explain the role of the Top-Ranking Sentences in each interface. In general, the techniques described in Chapter Three are used to extract and score Top-Ranking Sentences.

4.3.4.1 TRSPresentation Study

This study investigated the effectiveness of presenting a ranked list of Top-Ranking Sentences rather than a ranked list of documents. The Top-Ranking Sentences approach is compared against two interfaces that use traditional result presentation techniques (i.e., a ranked list of document titles, summaries and URLs). One experimental system ($S_{Baseline}$) directly presents the results from the underlying search engine and the other ($S_{TRSAbstract}$) uses the Top-Ranking Sentences as a document summary, presented below the document title in the same way as in Figure 3.2 (in Chapter Three). These two systems were compared against an experimental interface ($S_{TRSList}$). This interface, shown in Figure 4.1, consists of two main components: the Top-Ranking Sentences (that replace the traditional ranked document list) and a document pop-up window, which shows the subject more information about a particular document.

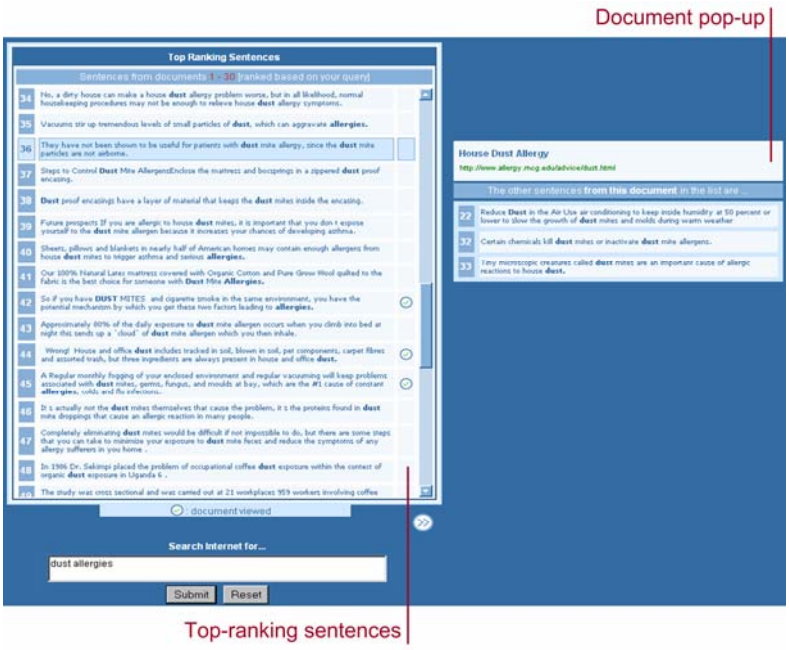

Figure 4.1. The experimental interface for the *TRSPresentation* study ($S_{TRSList}$).

The sentences are extracted and ranked using the techniques described in Chapter Three and presented in a list at the results interface. Initially there is no direct association between a

Top-Ranking Sentence in the list 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 clear. 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 the $S_{TRSList}$ interface the Top-Ranking Sentences drive searcher interaction whereas in the $S_{TRSAbstract}$ and $S_{Baseline}$ systems it is the titles, abstracts and URLs that encourage searchers to interact.

4.3.4.2 TRSFeedback Study

In this study the sentences are used to communicate the effects of relevance feedback decisions. For this purpose I developed two interfaces, one where the system endeavours to estimate relevance by mining searcher interaction ($S_{Implicit}$) and one where searchers had to explicitly mark information as relevant ($S_{Explicit}$). Unlike the $S_{TRSList}$ interface described in the previous study the order of the Top-Ranking Sentences in these experimental systems updates in the presence of relevance information. The two systems adapt to the context of the search by selecting additional query terms on the searcher's behalf based on relevance information provided during the examination of results. The only difference between the two systems is in how relevance information is conveyed. The $S_{Implicit}$ system makes the assumption that the viewing of a document summary (by moving the mouse pointer over its source document title) is an indication of searcher interest in the content of the summary. The $S_{Explicit}$ system requires searchers to explicitly indicate which results are relevant by clicking on checkboxes next to each document title. Figure 4.2 shows the interface to the $S_{Implicit}$ system.

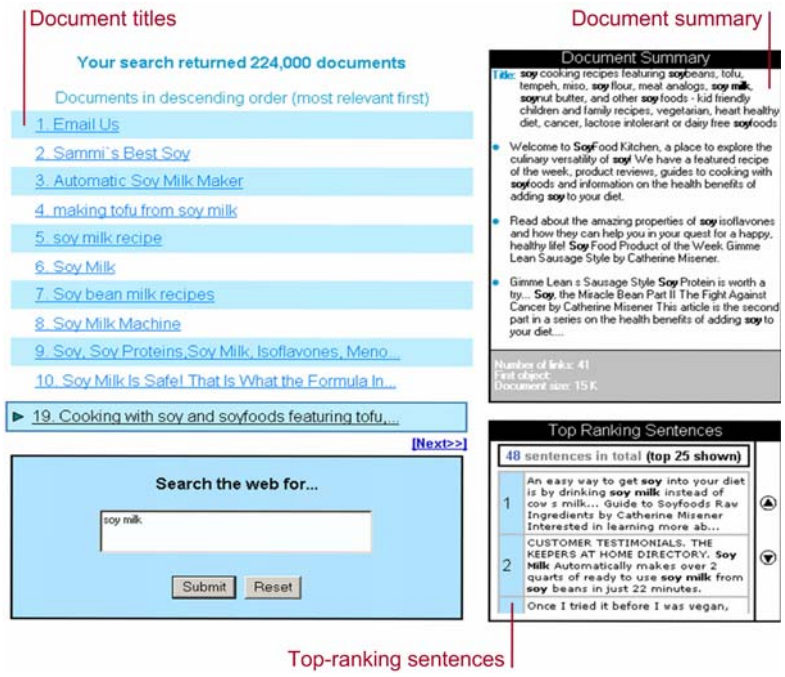

Figure 4.2. Experimental interface for the *TRSFeedback* study ($S_{Implicit}$).

After each relevance indication the summaries from the assessed relevant documents ($S_{Explicit}$) or assumed relevant documents ($S_{Implicit}$) are used to generate a ranked list of potential query modification terms using the *wpq* formula (Robertson, 1990). The top-ranked modification terms 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 based on term occurrence. The list of sentences is reordered after each relevance indication and due to the size of the window in which the sentences are displayed (shown in the bottom right-hand corner of Figure 4.2) only the top 25 sentences are displayed at any time. To make changes in the ordering of the list of sentences more noticeable, sentences from assessed summaries are removed from the list as the search progresses. The sentence reordering or the removal of Top-Ranking Sentences from the sentence list cannot be reversed by searchers.

4.3.4.3 TRSDocument Study

In a similar way to *TRSFeedback*, the experimental interface in this study uses implicit feedback techniques to gather relevance information and reorder a list of Top-Ranking Sentences. However, rather than communicating relevance feedback decisions the sentences (and the reordering) were used to facilitate access to retrieved documents. The experimental system ($S_{Feedback}$) automatically creates new search queries based on implicit feedback and is compared with a baseline summarisation system ($S_{Summarisation}$) used in White *et al.* (2003b)

and a system where the order of the sentence list is static and the query is assumed to be constant within an individual search iteration (S_{Static}). Figure 4.3 shows the interface used in the S_{Static} and $S_{Feedback}$ systems. The $S_{Summarisation}$ system uses the same interface without a list of Top-Ranking Sentences.

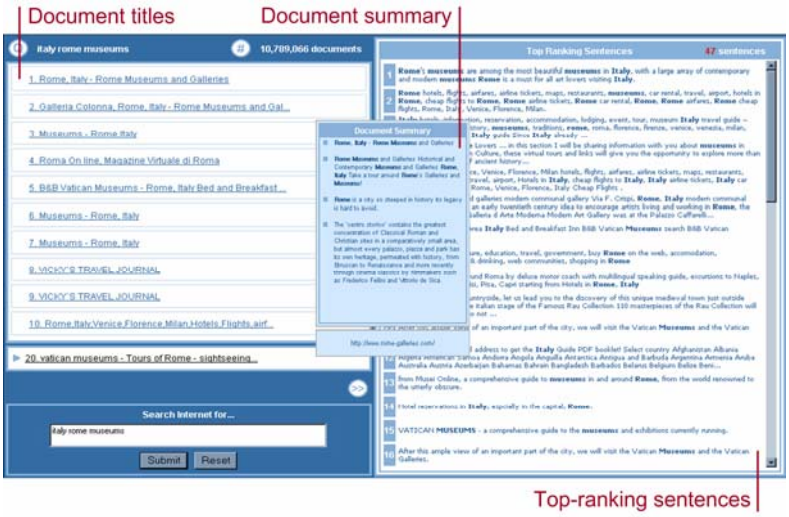

Figure 4.3. The experimental interface for the *TRSDocument* study (S_{Static} and $S_{Feedback}$).

As in the $S_{Implicit}$ system in *TRSFeedback*, the implicit feedback in this study is the evidence the searcher gives by viewing a document summary. To allow the system to better monitor this activity, the summary was moved to a pop-up window that appears when the mouse pointer hovers over a document title and disappears when it is removed from it. Once again the *wpq* method uses this evidence to select query modification terms on receipt of this relevance information. The ordering of the sentence list changes immediately when this information is provided and coincides with the presentation of the pop-up summary window. In $S_{Feedback}$ – as in the systems in *TRSFeedback* – sentences from relevant summaries are removed from the list to make the reordering more obvious and there is no option to reverse system decisions.

In *TRSFeedback* the system interprets every summary view as an indication of relevance. This led to problems of accidental 'mouseover', with searchers passing over document titles en route to those that interested them. In this study, the system implemented a timing mechanism that dealt with this problem and allowed me to base the implicit feedback on the length of time a searcher spent viewing a summary. Subjects conducted a timing task before they used each system, allowing the calculation of a relative viewing time for each subject and the $S_{Implicit}$ system to individuate its responses. This time was used for each subject as a determinant of whether a summary they viewed was relevant. From an analysis of all

subjects' viewing times from the timing task I found that they generally view relevant summaries for longer than non-relevant summaries (White *et al.*, 2002a). I use the viewing of document summaries as relevance indications since the system can easily detect which summaries are viewed and for how long.

4.3.4.4 Summary of Interfaces

All interfaces presented in this section encourage a deeper examination of search results and some use implicit feedback techniques to adapt the display in light of searcher interaction. In Table 4.2 I summarise the features of the systems created for each of the three user studies in three categories: *presentation* (i.e., how search results are presented) *summarisation* (i.e., how documents are summarised) and *feedback* (i.e., how relevance information is communicated).

Table 4.2
Features of experimental systems in the three user studies.

Feature	TRSPresentation			TRSFeedback		TRSDocument		
	$S_{Baseline}$	$S_{TRSAbstract}$	$S_{TRSList}$	$S_{Explicit}$	$S_{Implicit}$	$S_{Sum.}$	S_{Static}	$S_{Feed.}$
Presentation Method								
1. Top-Ranking Sentences			•	•	•		•	•
2. Ranked document list	•	•		•	•	•	•	•
Summarisation Method								
1. Chapter Three		•	•[α]	•	•	•	•	•
2. Search engine	•							
Feedback Method								
1. Explicit				•				
2. Implicit					•			•

[α] Although the $S_{TRSList}$ system does not present document summaries it uses the summarisation method described in Chapter Three to select Top-Ranking Sentences.

In this section we have described the experimental interfaces used in each of the three user studies. The systems within each study differ in ways necessary to test the experimental hypotheses. In the next section I describe the relationship between the three studies.

4.3.5 Inter-study Relationship

The studies all used Top-Ranking Sentences, but for a different purpose and to test different sets of hypotheses. Table 4.3 illustrates the main factors of each study.

Table 4.3
The main experimental factors in the three user studies.

	Study		
Factor	TRSPresentation	TRSFeedback	TRSDocument
Hypotheses	1. Top-Ranking Sentences are a viable alternative to Web document abstracts. 2. Top-Ranking Sentences increases awareness of result set content and is preferred by searchers. 3. Top-Ranking Sentences improve perceptions of task success, actual task success across all tasks.	1. Implicit relevance feedback is a viable substitute for explicit relevance feedback in Web retrieval.	1. The use of Top-Ranking Sentences encourages subjects to interact more fully with the retrieval results (i.e., documents) lead to more effective searching. 2. Implicit feedback improves subjects' perceptions of the system and leads to more effective interaction.
Factors measured	Search effectiveness, subject perceptions	Search effectiveness, subject perceptions	Search effectiveness, subject 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. Summarisation baseline 2. Summarisation/TRS 3. Summarisation/TRS/ Implicit Feedback
Subjects	18	16	24
Grouping	9 inexperienced 9 experienced	None	12 inexperienced 12 experienced
Age	Average = 23.80 yrs Range = 32 yrs (17:49)	Average = 24.75 yrs Range = 11 yrs	Average = 24.73 yrs Range = 33 yrs (16:49)
Internet Usage/week	Inexperienced = 4.2 hrs Experienced = 32.6 hrs	14 hrs	Inexperienced = 4.1 hrs Experienced = 29.8 hrs
Tasks	3 simulated work tasks (fact, decision and background)	4 each of Medical, Buying, Travel and Project	3 simulated work tasks (fact, decision and background)
Experimental design	Graeco-Latin square	Latin square	Latin square
Tasks per subject	3	4	3

Time per task	10 minutes	10 minutes	10 minutes
Data Collection	Five questionnaires (One demographic, three system and one final) Background logging	Five questionnaires (One demographic and four system) Background logging	Five questionnaires (One demographic, three system and one final) Background logging Semi-structured interviews

In *TRSPresentation* I encourage subjects 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 my techniques. Top-Ranking sentences are first introduced as a replacement for document lists; I then study the substitutability of implicit and explicit feedback using these sentences. I finish by using both documents and sentences in a more intricate form of implicit feedback, based on the proof of substitutability that *TRSFeedback* provided me with. Figure 4.4 shows the relationship between the three user studies.

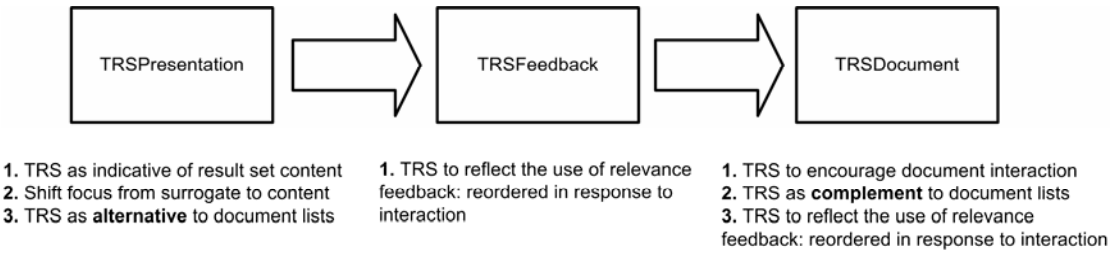

Figure 4.4. The relationship between the three user studies.

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 subjects. In the next section qualitative results from the studies are presented and the implications of them discussed.

4.4 Findings and Discussion

In this section I 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.* (2003a) (*TRSPresentation*), White *et al.* (2002b) (*TRSFeedback*) and White *et al.* (2002a) (*TRSDocument*). Since the studies were conducted with different subjects, on different systems, at different times, direct comparisons across studies is difficult. Therefore I focus mainly on subject opinions of the search process, the Top-Ranking Sentences and the implicit feedback used to reorder the sentences. The findings discussed motivate the systems developed in the remainder of this thesis.

4.4.1 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, I discuss the findings of the user studies in relation to this model.

The experimental systems described in this chapter 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, 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/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 subject has successfully completed their task. However, it is possible to minimise this anxiety by providing levels of support that help subjects 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 subjects 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 requires different levels of support from a search system. The systems described in this chapter 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 subjects 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". The Top-Ranking Sentences are a list of query-relevant document representations that may help subjects better understand their information need and begin conceptualising these needs to form search statements.

The systems presented in this chapter provide limited support for the *formulation* stage of the ISP. This assumes that there is a point of 'focus' (Kelly, 1963; Belkin, 1980; 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 personalised and search systems that fully support it help searchers construct query statements. In the systems described in this chapter 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. In the next section I discuss subject perceptions of the Top-Ranking Sentences.

4.4.2 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 subjects initially expressed concerns about the unfamiliarity of the interface. In this section I discuss subject perceptions of the TRS under three main section headings: task, popularity and usability.

4.4.2.1 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.

4.4.2.2 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 commercial 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*. Subjects 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* [13]). 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* I shifted the focus from document surrogates to the actual content of the document. In doing this, I found that the document titles were less useful as subject attention was drawn to the information resident inside documents. The experimental system used in *TRSPresentation* increased awareness of returned document set content, allowing subjects to make better decisions on the relevance of both the retrieved set of documents and documents individually.

4.4.2.3 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 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 4.4 shows the percentage differences with the experimental baselines ($S_{Baseline}$, $S_{TRSAbstract}$ and $S_{Summarisation}$) 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. All differences reported in the table were significant at $p < .05$.

[13] This was the only study where I measured the usefulness of the URL.

Table 4.4
Percentage difference between TRS 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

As can be seen from Table 4.4, 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 subjects were interacting in a way symptomatic of increased perusal with the retrieved document set. The shorter task completion times and increased number of tasks completed suggests that subjects were using their time more efficiently. In the next section I discuss the results obtained on the implementation of implicit feedback in the experimental systems.

4.4.3 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 expansion 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 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* I 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 criticised; searchers will view non-relevant 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 I felt that this assumption was fair enough to allow an initial investigation into the use of implicit feedback. In *TRSDocument* I 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. More complex and effective techniques based on these findings are described in later chapters of this thesis.

Despite their positive comments, subjects had two reservations about how system decisions based on implicit feedback were communicated. Firstly, since 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. Secondly, the Top-Ranking Sentences only contained sentences from Web pages for which the subject had not already viewed a summary. If the subject 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. In White (2004) I proposed the use of *ScrollTiles* to communicate the effects of the sentence reordering using a familiar interface component, the scrollbar. The approach represented sentences as tiles on the scrollbar and re-coloured the tiles to represent changes in the ordering. A pilot study was conducted that involved nine experimental subjects and compared systems that re-coloured a representation of the sentences imposed on the scrollbar with one that reordered the actual sentences. The ScrollTiles were shown to be more effective for conveying reordering decisions than the sentence updating. However, they are not used in any further interfaces described in this thesis as I tried to limit the number of new interface components to only those necessary to test experimental hypotheses.

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 were useful in the development of search interfaces described later in this thesis.

4.5 Summary

Ranking documents is potentially 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. The content-driven approach extracts, ranks and presents the content of the returned set, blurring inter-document boundaries and encouraging information seeking based on the potentially relevant document content.

In this chapter I have discussed the results of three studies to test the effectiveness of content-driven information seeking. The implicit feedback frameworks proposed in this thesis rely on searcher interaction with the retrieved information as evidence of what information is relevant. The studies presented in this chapter show that the interfaces developed are liked by subjects and can lead to more effective information seeking. This was a promising finding for the development of search systems developed later in this thesis. The studies have also highlighted problems in the use of these interfaces that are addressed in later systems. In Chapter Five I present an overview of a search interface that uses titles, summaries and Top-Ranking Sentences and other document representations to facilitate access to potentially relevant information.

Chapter 5

Representations and the Search Interface

5.1 Introduction

In this chapter I describe the document representations presented at the search interface and used by the implicit feedback frameworks described in the thesis. These representations are typically sentence-based and created by the search system at retrieval time. As they are small, interaction with document representations is potentially more focused than with the full-text of documents and since they are numerous, can generate an increased quantity of evidence for the implicit feedback frameworks. In Chapter Four document representations were used to encourage searchers to interact more with the results of their search. Through presenting multiple representations of the same document it is possible for searchers to directly indicate which document components (e.g., sentences, summaries, and titles) are relevant. Traditional RF techniques rely on searcher feedback about the relevance of whole documents; this can be unreliable as documents can contain irrelevant parts. The principle of *polyrepresentation* (Ingwersen, 1994) suggests that IR systems should provide and use different cognitive structures during acts of communication to reduce the uncertainty associated with interactive IR. The techniques I describe implement one aspect of a polyrepresentative approach; the use of multiple document representations.

The chapter also presents the generic design of a search interface that combines document representations in an interactive context. The document representations and interface presentation techniques are described in the remainder of this chapter.

5.2 Document representations

IR systems were originally designed for the retrieval of documents from homogeneous corpora, such as newspaper collections or library index cards. Document surrogates, such as

titles and abstracts, were usually created by experts, such as librarians or professional cataloguers. The growth in size, dynamism and heterogeneity of the collections being searched led to the development of automated representation techniques and a reduction in the quality of the surrogates created. However, work by Landow (1987) and Furnas (1997) has shown the importance of the information that searchers use when deciding which documents to download and view. If the quality of document representations has decreased, then one possible solution is to increase the quantity of information available to view. That is, provide searchers with more information to make search decisions.

In my approach the most relevant documents in the retrieved set are represented by a variety of document representations. The principle of polyrepresentation (Ingwersen, 1994) suggests that different cognitive structures should be offered to searchers and used by them during their interaction with an IR system. The cognitive structures around which polyrepresentation is based are manifestations of human cognition, reflection or ideas. In IR they are typically transformations generated by a variety of human actors with a variety of different *cognitive origins*. The author's text, including titles and the full-text are representations of cognitive structures intended to be communicated. However, these portions of text have different *functional origins*. That is, they have the same cognitive origin but were created in a different way or for a different purpose.

In Chapter Four experimental search interfaces were presented that used different representations of the top-ranked documents. In those studies Top-Ranking Sentences, titles and document summaries were used to represent their source documents and facilitate effective information access. In this chapter, three further representations are used: summary sentences, summary sentences in document context and the full-text of the document. These representations describe the document in different ways. The full-text is only the textual content of the document; all other document features, such as images and document structure, are ignored since they cannot be used by the sentence selection methods described in Chapter Three.

The sentence-based representations (i.e., Top-Ranking Sentences, document summaries, summary sentences and sentences in context) have different functional origins and the same cognitive origins (different from the author of the source document). These representations are created using algorithms devised by the system designers and are selected based on queries submitted by a searcher, both cognitive agents. Offering searchers different representations of the same document at the search interface is one aspect of polyrepresentation. However, the basis of polyrepresentation is the use of the overlap

between these representations to reduce uncertainty. The theory has been implemented across networks of citations (Larsen and Ingwersen, 2002), where those who cite documents have unique cognitive structures. The interface described in this chapter use many document representations to implement one aspect of a polyrepresentative approach that aim to reduce the uncertainty associated with gathering implicit feedback. In this section I introduce each of the representations and explain their role in the search interface.

5.2.1 Top-Ranking Sentences

Top-Ranking Sentences were introduced in Chapters Three and Four as a means of facilitating access with retrieved information. The results of the user studies in Chapter Four demonstrated the usefulness of presenting sentences in a list, ranked independently of their source documents. The interfaces described in this chapter use these sentences in the same way. Ingwersen (1994) suggests that paragraphs are the smallest semantically confined unit of a document that can effectively be used in any application of polyrepresentative principles. Paragraphs have been used as passage-level evidence for the indexing and subsequent retrieval of documents (Salton *et al.*, 1993; Callan, 1994). In the search interfaces I create, the Top-Ranking Sentences provide a starting point from which searchers can access potentially useful information. The sentences may contain the information necessary to satisfy their information need, or may provide a means through which searchers can access relevant documents.

5.2.2 Document Title

This is the title of the document, as assigned by the author. Titles are typically short and include terms that express the main themes of a document. On the Web, the corpus for the user studies described in this thesis, authors assign document titles and the extent to which they are indicative of current document content can vary.

5.2.3 Document Summary

A document summary contains the four Top-Ranking Sentences for that document. The summary is based on the query submitted by the searcher and is created in real-time, when a query is submitted, using the best Top-Ranking Sentences selected by the approach given in Chapter Three. Figure 5.1 shows an example summary produced by the Google Web search engine. This summary is typically composed of a series of sentence fragments that could contain the query terms, separated by ellipses.

Modern **Information Retrieval**
(back cover). Modern **Information Retrieval**. ... Inc. You can order this book on-line with a secure form, or search other titles from A/W about **Information Retrieval**. ...
www.sims.berkeley.edu/~hearst/irbook/ - 9k - Cached - Similar pages

Figure 5.1. Document abstract from Web search engine for query 'information retrieval'.

Figure 5.2 shows the summary generated by combining the same document's four best Top-Ranking Sentences. The summary window on the right of the figure appears immediately or after a short time delay when the searcher hovers over the document title.

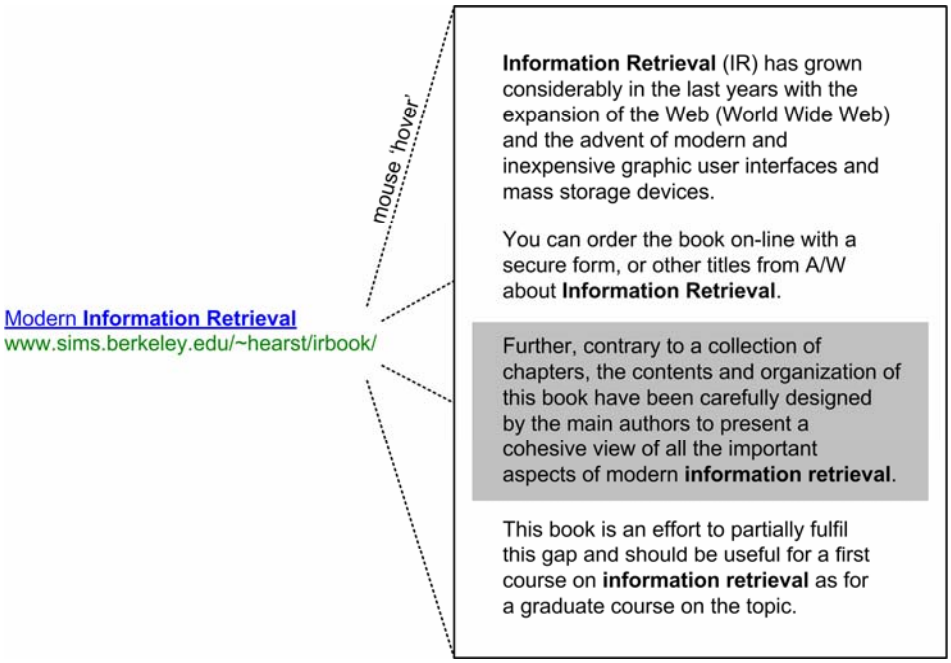

Figure 5.2. Document summary from the best four Top-Ranking Sentences for query 'information retrieval'.

The difference in the content and quality of the summaries between the two summary generation approaches is significant. The summaries created by combining the best Top-Ranking Sentences are semantically richer and may allow more accurate relevance assessments than standard search engine summaries (White *et al.*, 2003b).

5.2.4 Summary Sentence

Each sentence in the summary of the document is considered a representation of the source document. Allowing relevance assessments at the sentence level allows for more precise assessments of what information meets searchers needs. In Figure 5.2 the third summary sentence is highlighted.

5.2.5 Sentence in Context

A summary sentence in the context in which it occurs in the document (i.e., preceding and following sentence from the source document) is also available for searchers to view. This can be of particular use when a sentence is *anaphoric* i.e., refers back to a previous sentence in the document or *cataphoric* i.e., refers forward to a forthcoming sentence in the document. For example, if there are the two sentences: "Alexander Graham Bell invented the telephone. He emigrated to Canada when he was just 23". The pronoun 'he' in the latter sentence is referent to the "Alexander Graham Bell" in the former sentence. This is an anaphoric reference and can be problematic if the latter sentence is shown without the first. Presenting the latter sentence in the original document context can contribute to the resolution of such problems.

In Figure 5.3 the highlighted sentence in Figure 5.2 is shown in the context in which it occurs in the source document. The sentence in context appears directly next to the sentence in summary to make the association between the two representations more clear. In the 'Sentence in Context' window on the right of Figure 5.3, the summary sentence is highlighted and the preceding and following sentences are also shown to the searcher.

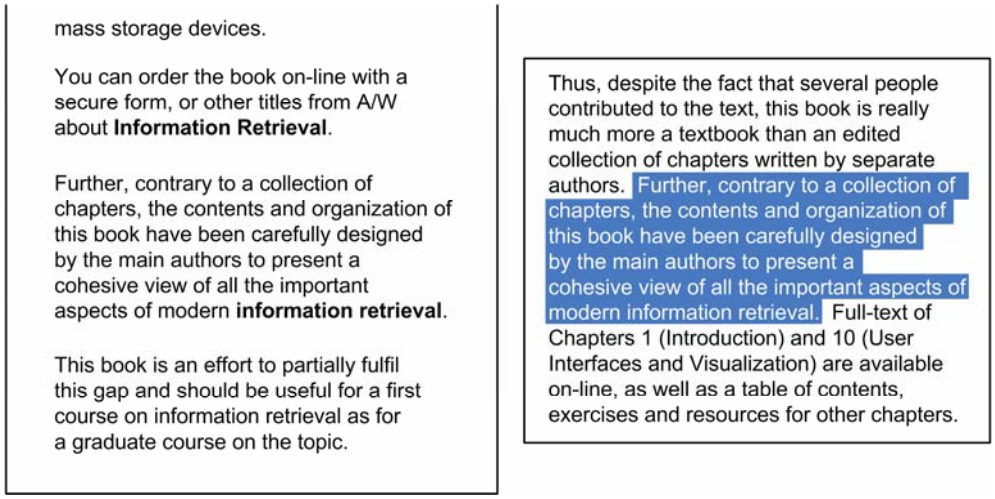

Figure 5.3. Summary sentence in document context.

The sentences in context are created immediately after the retrieved documents have been summarised (i.e., after query submission and before result presentation). Figure 5.4 shows the process involved to create the sentence in context for each sentence in the document summary. First the Top-Ranking Sentences are selected from the source document, and the sentences that comprise the summary are passed to the context generation component. Each sentence has a unique identifier, based on its position in the document. The context

generation component then locates the sentence that immediately precedes and immediately follows the summary sentence. For example, in Figure 5.4 sentence s_3 is a summary sentence and sentences s_2 and sentence s_4 form the context for s_3.

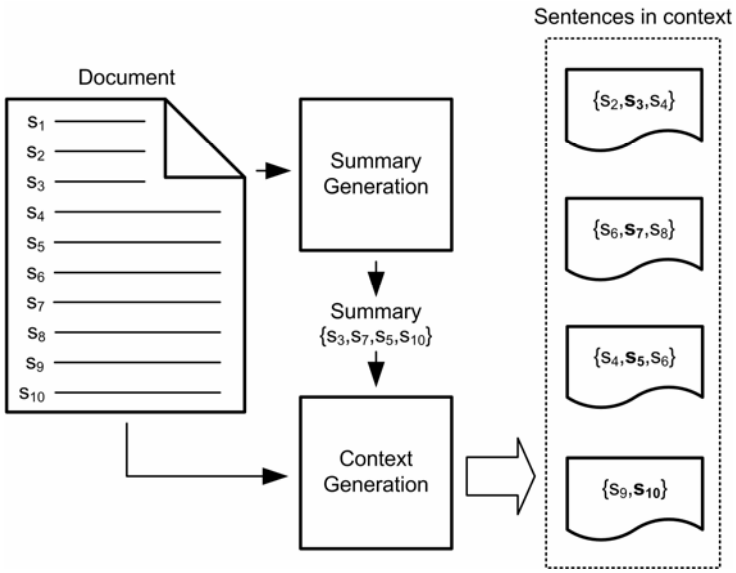

Figure 5.4. Creation of sentence in document context.

If a sentence is the last sentence in a document (as with s_{10} in Figure 5.4) only the sentence before is used to compose the sentence in context. Since s_{10} is the last sentence, the context will only comprise s_{10} and the prior sentence s_9. The same is true for the first sentence, except that the only the sentence directly following it is used to comprise the context.

5.2.6 Document (Full-text)

The full-text is the document, as created by the author. The full-text of the document is the source of the sentences used to create the document representations. Monitoring searcher interaction with documents is problematic as it can be difficult to determine exactly what part of the document, if any, searchers regard as relevant. Using all terms from documents searchers view may adversely affect the retrieval performance of the term selection parts of the model (Salton *et al.*, 1993), especially if the document is actually irrelevant. Therefore, the document full-text is not used directly in any of the implicit feedback frameworks described in this thesis. However, the set of terms extracted from the set of most relevant documents forms the vocabulary or term space used by the implicit feedback frameworks described in later chapters.

5.2.7 Overview of Representations

There is redundancy in the representations that searchers interact with. A single top-ranking sentence may appear in five of the six document representations: the Top-Ranking Sentences list, the document summary, a summary sentence, a sentence in context and the source document. Searcher interaction with the same sentence in a number of representations provides more evidence for the relevance of representations.

Different types of representation vary in length, and can hence be regarded as being more or less *indicative* of the content of the document (Barry, 1998). For example, a top-ranking sentence is less indicative than a query-biased document summary (typically composed of four sentences) as it contains less information about the content of the document. The *length hypothesis* (Marcus *et al.*, 1978) suggests that the quality of a representation is directly proportional to its length. The validity of this hypothesis had been supported by previous work (Weis and Katter, 1967; Hagerty, 1967). However, the hypothesis has been criticised for failing to consider the quality or nature of a representation (Janes, 1991). For example, a document title is typically short but is assigned by the author, and may capture the key concepts in a document. The heuristic-based implicit feedback framework described in Chapter Six uses the length hypothesis to assign an indicativity weight to the representations.

An alternative approach is to assume that representations that are more indicative of their source documents contribute more to the refinement of query statements. Janes (1991) views the length hypothesis as superficial and perhaps more suited for heuristic-based approaches. The probabilistic implicit feedback framework presented in Chapter Seven does not use representation length as a measure of representation quality. Instead, it gives more weight to representations with higher quality content. To do this, it constructs an *indicativity index* (White *et al.*, 2004b) measured based on the terms that co-occur between the representation and the document. Representations that are highly indicative of the source document are regarded as high quality. Some representations of each document are fixed in content, i.e., the title and full-text of the document, whereas other representations, such as the summary, are dependent on the query and hence variable in content. The document title and the full-text are created by the author and are not query dependent.

In the next section I describe the search interface that combines the document representations for the presentation of search results.

5.3 Search Interface

The search interface presents a variety of document representations to the searcher. These *content-rich* search interfaces present more information from retrieved documents than standard search engine interfaces. Through their interaction searchers can control which representations are shown on the interface at any one time. A schematic of the interface is shown in Figure 5.5. The 'Summary', 'Sentence in Context' and 'Document full-text' all become the active window – displayed in front of the other information – when the searcher requests them. The default display is the list of Top-Ranking Sentences and the list of document titles. The list of Top-Ranking Sentences can contain around 60 sentences from the most relevant Web documents.

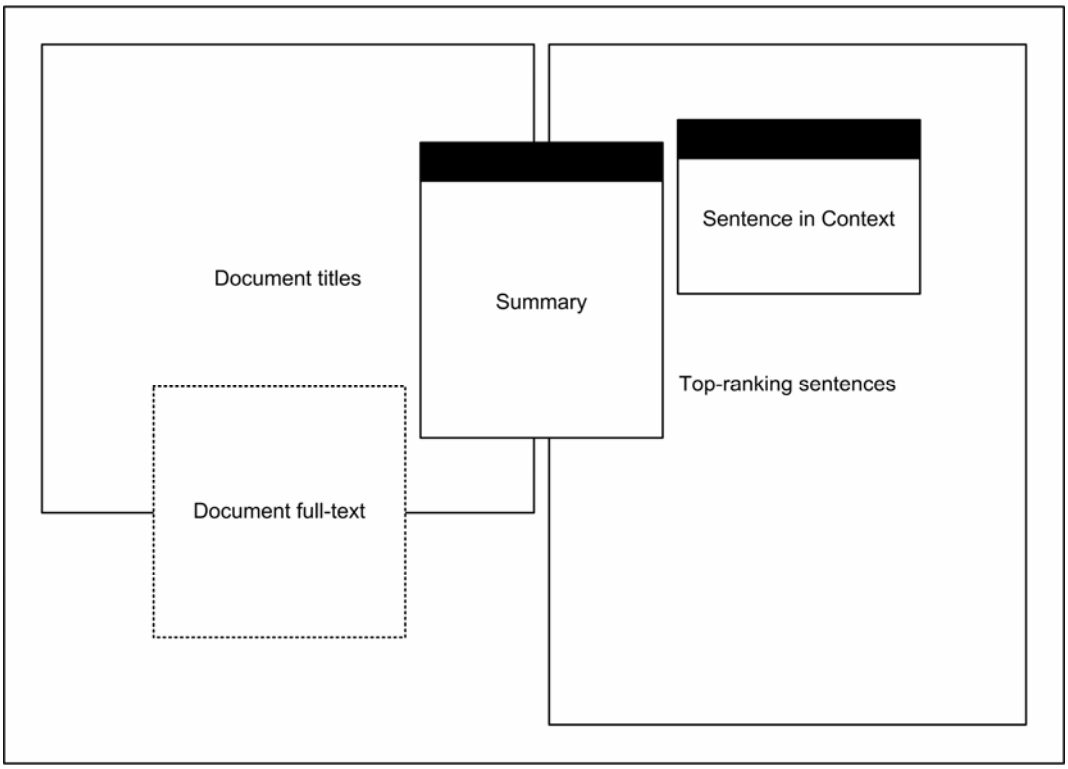

Figure 5.5. Schematic of the search interface.

This style of interface was chosen since it allows the search system to closely monitor what document representations searchers may be viewing at any given time. This allows implicit feedback frameworks that use interaction with these interfaces to make potentially more accurate inferences about searcher interests. Searchers can view the title of a top-ranking sentence's source document simply by interacting with the sentence. Should the title fall outside the first 10 documents then a small window below the list of document titles updates to show the title (as a clickable hyperlink) and in some systems the URL. An example of this window is given in Figure 5.6.

Figure 5.6. Document title pop-up for documents outside the top ten retrieved.

Searchers can interact with the hyperlink in this window in the same way as with any title in the first 10 retrieved documents. That is, they can click the text to visit the document or hover over the title to see a summary of the document. Figure 5.7 shows an experimental interface used in Pilot Test 1, described in Chapter Nine, which implements these concepts.

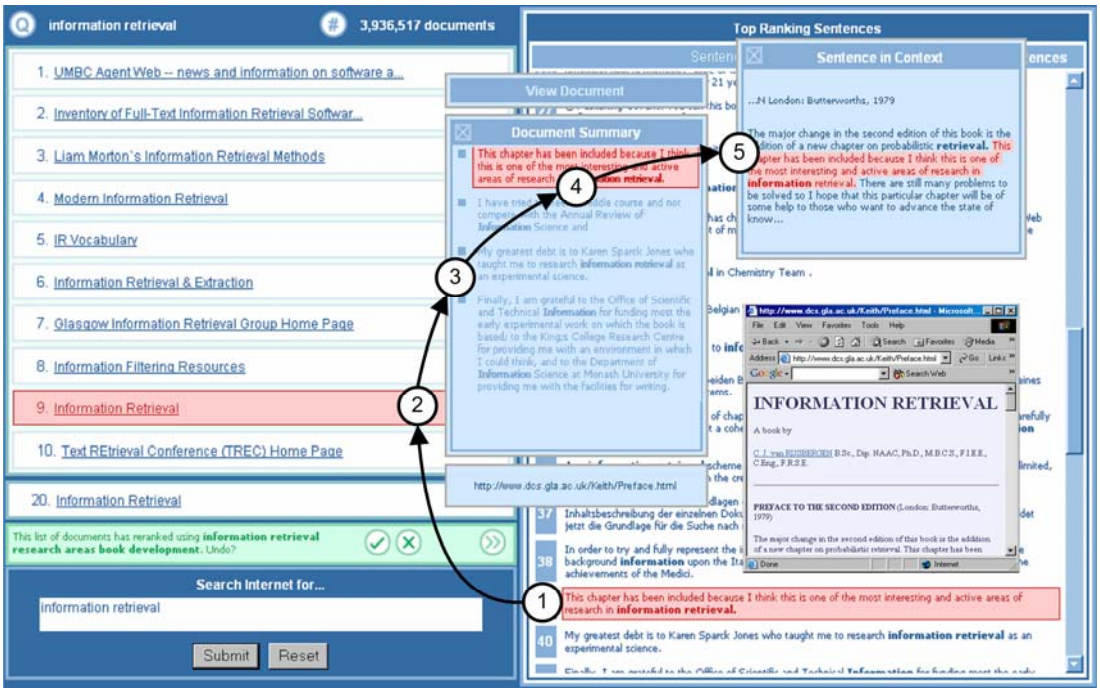

Figure 5.7. Experimental search interface in Pilot Test 1 (Chapter Nine).

The effectiveness of top-ranking sentence-based interfaces to statically structure information spaces has already been demonstrated (Tombros *et al.*, 2003a; 2003b). In these studies Top-Ranking Sentences were clustered to create personalised search spaces that made interaction more effective. The implicit feedback frameworks described in this thesis modify the query and estimate changes in the information needs of searchers. Adaptive views of the information space can support the developing nature of information needs (Campbell, 1999). The frameworks restructure or recreate the search results at each query iteration to bring potentially relevant results to the attention of the searcher. The mechanisms behind the interface proposed in this section use searcher interaction to formulate a query that represents their information need and dynamically restructure or recreate the search results based on the predicted extent of any changes in this need.

Interacting in a certain way with each representation suggests another representation for that document. For each document it is possible to follow a path between its representations. These are called *relevance paths* since the further a searcher travels along a path the more evidence there is on the relevance of the path's resident information. Searchers are guided along the relevance path by their interaction and the search system. In the next section these paths are described.

5.4 Relevance Paths

There are many applications of paths in IR (Pirolli and Card, 1995; Campbell and Van Rijsbergen, 1996; Chalmers *et al.*, 1998). The Ostensive Model (Campbell and Van Rijsbergen, 1996) uses paths between documents or document representations to build a context for the search and choose appropriate terms to form a new query. Information foraging theory (Pirolli and Card, 1995) assumes users are driven by the to click hyperlinks based on proximal cues given by their surrounding text. The path model (Chalmers *et al.*, 1998) uses each individuals' ongoing history of ratings or choices to choose similar pages.

These applications all consider paths between documents e.g., clicking a hyperlink resident in one document to get to another document. However, the relevance paths I propose form *between document representations*. The paths provide searchers with progressively more information from the best documents to help them choose new query words and select what new information to view. The further along a path a searcher travels (i.e., the more representations in a path they view) the more relevant the information in the path is assumed to be. The order in which certain types of representation are available in a relevance path is dictated by the interface. Searchers are guided along the path by their interaction with the search interface. If they interact with the Top-Ranking Sentences the system highlights the title of the source document. If they hover over a document title for a short time the summary of that document appears in a small, moveable window in front of the other information. Clicking arrows next to sentences in that summary shows the sentences in the context they occur in the source document.

The paths can vary in length from one to six representations long, and searchers can access the full-text of the document from any step in the path by clicking on the text of the document representation. Since searchers can take many routes between representations for each document, there may be many *potential* relevance paths. Relevance paths can start from Top-Ranking Sentences or document titles. Certain aspects of the path order are fixed e.g., the searcher must view a summary sentence before visiting that sentence in context. The full-text

of the document is accessible from all representations. That is, a searcher can click on all representations and access the source document. There are 54 potential relevance paths for each document. In Figure 5.8 I show a possible relevance path route for a single document, at each step a representation is viewed (shown in darker font).

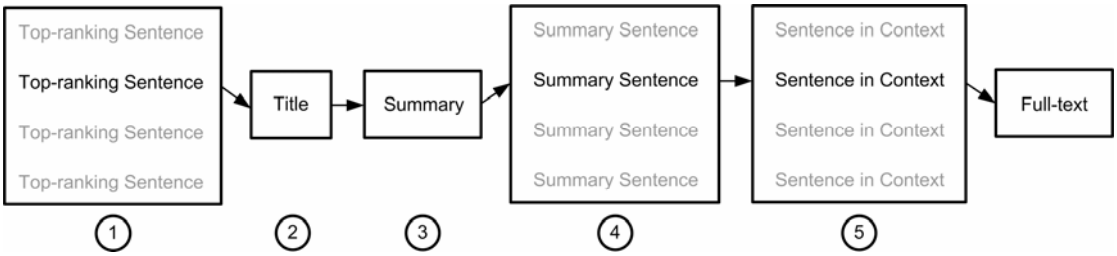

Figure 5.8. Possible relevance path route (numbers correspond to Figure 5.7).

In Figure 5.8 a top-ranking sentence is viewed, followed by the title of the document, the summary for that document, a sentence in that summary and in context, followed by the full-text of the document. There are six steps in this relevance path. In Figure 5.9 this relevance path is shown on the interface schematic. To follow this path a searcher would have to interact with each of the representations on the path. The full-text of the path's source document is eventually accessed in this instance from the sentence in context.

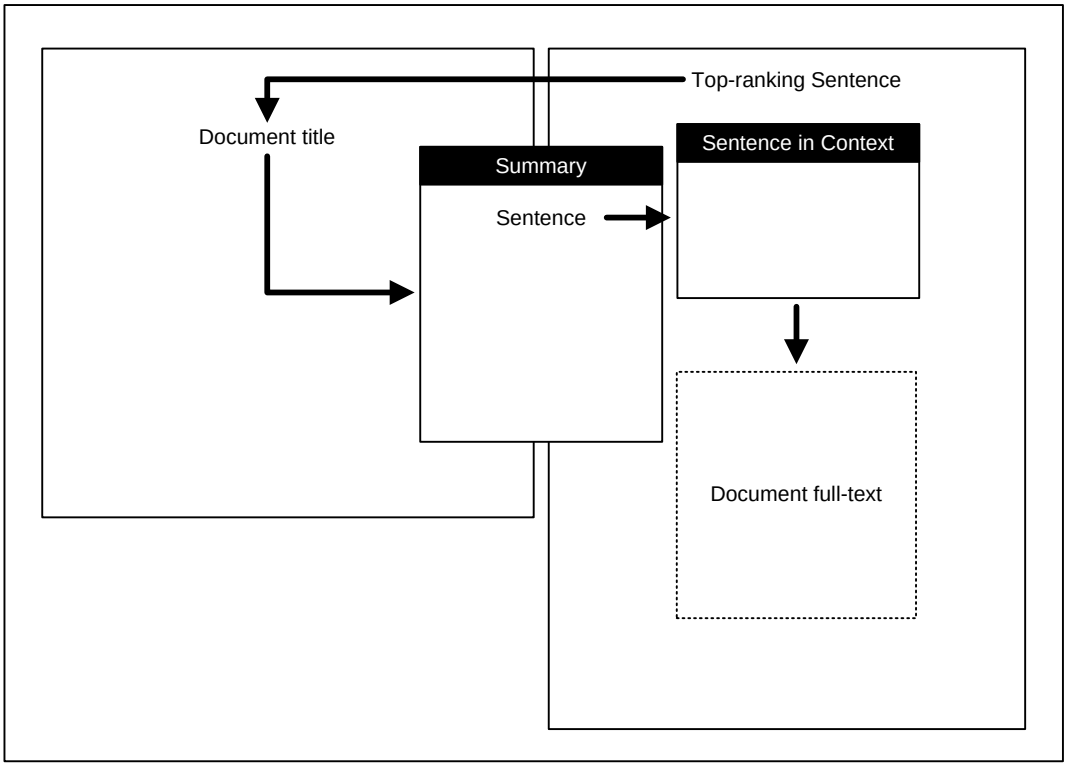

Figure 5.9. Possible relevance path on interface schematic.

As a searcher moves along the relevance path they move from assessing document representations in relation to other representations (i.e., Top-Ranking Sentences, titles) to a deeper examination of representations in their resident context (i.e., summaries, sentences in context). That is, as a searcher traverses a relevance path, their interaction with top-ranked documents becomes more focused. To the searcher, the path represents a desire to find out more information about a document or to find the information they require to satisfy their needs. To the implicit feedback framework operating behind the search interface, each relevance path is a source of evidence that allows it to build a body of relevance and make decisions on the searcher's behalf. Showing searchers progressively more information about a document to assist relevance assessments has already been used in related work (Zellweger *et al.*, 2000; Paek *et al.*, 2004).

5.5 Summary

In this chapter I have described the document representations presented to searchers at the interfaces described in this thesis. These representations allow searchers to view and assess the relevance of information at the results interface rather than visiting documents and locating the information inside them.

Document representations are linked at the interface by relevance paths that guide searcher interaction. The further along a relevance path a searcher travels, the more relevant the information in the path is assumed to be. These paths are included in the content-rich interfaces described in this chapter and aim to encourage searchers to interact with the retrieved information in a structured way, generating more evidence for the implicit feedback frameworks that use this as evidence of searcher interests. In forthcoming chapters the frameworks that utilise this interaction are presented.