Question answering, relevance feedback and summarisation: TREC-9 interactive track report

Neill Alexander, Craig Brown, Joemon Jose, Ian Ruthven[1] and Anastasios Tombros
Department of Computing Science
University of Glasgow, Glasgow, G12 8QQ. Scotland
alexannw,brownca,jj,igr,tombrosa@dcs.gla.ac.uk

Abstract

In this paper we report on the effectiveness of query-biased summaries for a question-answering task. Our summarisation system presents searchers with short summaries of documents, composed of a series of highly matching sentences extracted from the documents. These summaries are also used as evidence for a query expansion algorithm to test the use of summaries as evidence for interactive and automatic query expansion.

1. Introduction

The main focus of Glasgow's Interactive Track study was to investigate the use of summaries in interactive searching.

Our experiments used a form of the query-biasing summarisation technique proposed by Tombros and Sanderson [TS98], to create short document summaries that are tailored to the user's query. These summaries highlight the main points of the document that pertain to the query. The summaries are based on highly matching sentences, allowing users to view the context in which query terms are used within the document.

We hypothesised that this form of summarisation would be particularly effective for the time-limited, query-answering task of the interactive track, in that summaries would allow users to filter out non-relevant documents more effectively and target potentially relevant documents more quickly than either title alone or the full text of the documents.

In addition, we investigated the use of summaries for relevance feedback (RF): by using the content of the summaries, rather than the full-text of the documents, to generate query expansion terms.

Our experiments indicate that although RF is generally not considered helpful, summaries can provide a popular and useful aid to finding relevant information.

The paper is structured as follows: in section 2 we describe the system we used in these experiments including details of the summarisation process, in section 3 we describe the interface, in section 4 we give details of the experimental subjects and in section 5 we analyse the results. We conclude in section 6.

[1] Corresponding author.

2. System

In section 2.1 we outline the main components of our system and in section 2.2 we describe how the summaries are created.

2.1 System architecture

Our experimental system was composed of three units:

1. *retrieval system*. The retrieval system (SMART) performs an initial query run using the query terms passed from the interface. The list of retrieved document identifiers and document titles are passed to the interface for display.
2. *summariser*. For each retrieved document, the summariser (described in section 2.2) generates a query-biased summary, which is passed to the interface on demand.
3. *interface*. The interface displays the retrieved document identifiers, document titles and summary. The overall look and feel of the interface is described in section 3.1. The interface is also responsible for logging user interaction and generating query expansion terms (described in section 3.2).

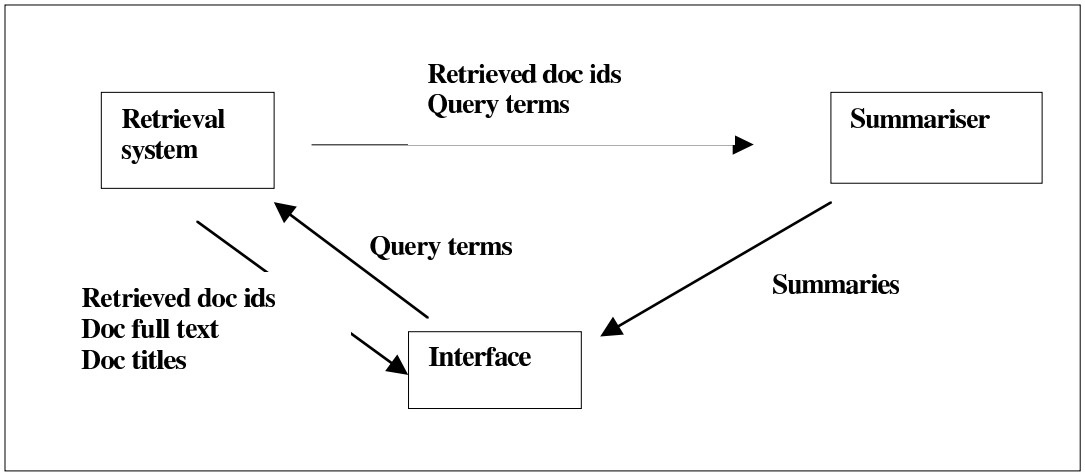

Figure 1: System architecture

2.2 Summariser

A document summary conventionally refers to a condensed version of a document that succinctly presents the main points of the original document. *Query-biased* summarisation methods generate summaries in the context of an information need expressed as a query by a user. Such methods aim to identify and present to the user individual parts of the text that are more focused towards this particular information need than a generic, non-query-sensitive summary. In this way summaries can serve an *indicative* function, providing a preview format to support relevance assessments on the full text of documents [RSZ71].

Query-biased text summarisation is an emerging area of summarisation research that had not been addressed until recently. Tombros and Sanderson looked into the application of such methods in information retrieval, evaluating the indicative function of the summaries [TS98]. Their study showed that users were better able to identify relevant documents when using the summaries than when using the first few sentences of a document. Recently the TIPSTER funded SUMMAC project [MHKHOFCS98] provided a framework for the evaluation of different types of summarisation systems. As part of that project, a number of query-biased summarisation systems were evaluated by measuring their ability to help users identify documents relevant to a query.

The summaries generated by our system were indicative and query-biased, aiming to provide users working on an interactive IR system with information on the relevance of documents retrieved in response to their query. The

system is based on a number of sentence extraction methods that utilise information both from the documents of the collection and from the queries submitted, and is a simplified version of the system described in [TS98].

Each document that was retrieved in response to a specific query was passed through the summarisation system, and as a result a score for each sentence of each document was computed. This score represents the sentence's importance for inclusion in the document's summary. Scores are assigned to sentences by examining the structural organisation of each document, and by utilising the inverse document frequency (IDF) weights assigned to each term. Information from the structural organisation of the documents was utilised in two ways. Terms occurring in the title section of a document were assigned a positive weight (title score) in order to reflect the fact that headlines of news articles tend to reveal the major subject of the article. In addition, a positive ordinal weight was assigned to the first two sentences of each article, capturing the informativeness of the leading text of news articles. The IDF weights of the terms in the collection were used as a source of evidence of attributing an overall measure of importance for each sentence of the source documents. In order to establish such a measure, the sum of the IDF weights of all the terms comprising a sentence was divided by the total number of terms in that sentence. In that way an importance score was attributed to each sentence in the collection.

In addition to the scores assigned to sentences, information from the query submitted by the user was also employed in order to compute the final score for each sentence. A query score was thus computed, intended to represent the distribution of query words in a sentence. The rationale for this choice was that, by allowing users to see the context in which the query terms occurred, they could better judge the relevance of a document to the query. The computation of that score was based on the distribution of query terms in each sentence. This was based on the belief that the larger the number of query terms in a sentence, the more likely that sentence conveyed a significant amount of the information need expressed in the query. The actual measure of significance of a sentence in relation to a specific query, was derived by dividing the square of the number of query terms included in that sentence by the total number of the terms comprising the query.

The final score for each sentence is calculated by summing the partial scores discussed above. The summary for each document is then generated by selecting the top-scoring sentences, and outputting them in the order in which they appear in the original document. Summary length was defined to be 20% of the document's length, up to a maximum of 6 sentences. Such a value seems to be in general agreement with suggestions made by [Ed64, BMR95].

Figure 3 shows the summary produced from the document in Figure 2, retrieved in response to the query '*America national parks redwood trees*'. In Figure 2 bold type marks those sentences that were extracted to form the summary.

<DOC>
<DOCNO> SJMN91-06312178 </DOCNO>
<ACCESS> 06312178 </ACCESS>
<DESCRIPT> CALIFORNIA; TREE; PARK; US </DESCRIPT>
<LEADPARA> **California's majestic redwood parks may be ceded to the federal government under a cost-cutting proposal under study by state Parks and Recreation Department officials, the officials said Wednesday.**; The three parks -- Jedediah Smith Redwoods State Park and Del Norte Coast Redwoods State Park in Del Norte County, and Prairie Creek Redwoods State Park in Humboldt County -- are the crown jewels of the state park system and home to 2,000-year-old redwoods, among the oldest living things on Earth. </LEADPARA>
<SECTION> California News </SECTION>
<HEADLINE> STATE MAY CEDE ITS 3 REDWOOD PARKS TO U.S. </HEADLINE>
<TEXT> The proposal emerged from a review ordered by state Parks and Recreation Director Henry R. Agonia after the Wilson administration sent a directive to, state agencies asking them to identify budget cuts. The state is facing a staggering $2 billion deficit in this year's $55.7 billion budget.; The prospect of transferring California's redwood parks to the National Park Service drew praise and criticism from environmentalists and park rangers Wednesday as word spread.; "I'm strongly against it," state parks Superintendent Bill Beap said in a telephone interview from Eureka. "These are the prime jewels of the state park system."; But the proposal was welcomed by Sierra Club officials, who called it "a splendid idea." **Edgar Wayburn, the club's vice**

president for conservation, noted that the neighboring National Redwood Park's boundaries touch all three state parks. </TEXT>
<BYLINE> Los Angeles Times </BYLINE>

Figure 2: Sample document

<SUMMARY>
<TITLE> STATE MAY CEDE ITS 3 REDWOOD PARKS TO U.S.
</TITLE>
<DOCID> SJMN91-06312178 </DOCID>
<LI>California's majestic redwood parks may be ceded to the federal government under a cost-cutting proposal under study by state Parks and Recreation Department officials, the officials said Wednesday. </LI>
<LI> " Edgar Wayburn, the club's vice president for conservation, noted that the neighboring National Redwood Park's boundaries touch all three state parks. </LI>
</SUMMARY>

Figure 3: Summary produced from document SJMN91-06312178

3. Interface

3.1 Look and feel

The interface consists of four main components: list of 20 retrieved document titles and associated check boxes for marking documents relevant, summary display, query box and suggested expansion terms.

A summary was generated on demand each time the user moved the mouse pointer over a document title. The complete document could be viewed, in a separate window, by clicking on the document title.

Query expansion was both automatic (the top 6 expansion terms were automatically added to the query when the user requested more documents), and interactive. The user could request suggestions for new query terms by clicking the 'Get More Terms' button, the suggested terms appearing in the 'More Terms' box on the bottom left of the screen. Each query, and suggested, term was displayed in this box; check-boxes were used to add/delete terms from the query.

3.2 Relevance feedback

To assist users in modifying their queries, a query expansion facility was offered. Expansion terms were selected from the summaries of marked relevant documents using Porter's term weighting function [PG88]. This is shown in Equation 1, where r equals the number of relevant summaries containing a term, R is the number of relevant documents, n is the number of documents in the collection containing the term, and N is the number of documents in the collection.

$$Porter = r/R - n/N$$

Equation 1: Porter term weighting function

This use of Porter's term weighting function differs from other applications in that r is based on the terms appearance in a summary of a relevant document, rather than the whole document. The basis for this is that a query-biased summary of a document may be a better source of relevant terms than the full-text of the document. This may be especially true for long documents that cover many topics.

For this experiment we used only the summary to determine possible expansion terms, even if the user had viewed the full-text of the document. A natural enhancement of this may be to vary the source of the expansion terms according to the representation(s) of the document viewed by the user. For example if the user has only viewed a summary before making an assessment, then the summary should be used for feedback, if the user has viewed the full text of the document, then the full-text may be a better source of evidence for feedback.

This method of calculating relevance weights has previously been shown to give good results [PG88, Efth95] for automatic and interactive query expansion. The expansion terms produced in our case, however, were not always useful. This was for two reasons: low numbers of relevant documents, and the processing of summaries for relevance weighting.

The searchers tended to find relatively low numbers of relevant documents. If a searcher only found one relevant document in the first display of documents (as many did) then the relevance weighting prioritised those terms that only appeared in the relevant document. In this case, not only were these terms not useful for retrieving further relevant documents but occasionally the terms turned out to be spelling errors in the original documents, e.g. '*armioffici*', '*withth*', and '*sovietunion*'.

A flaw in our preparation of the summaries for relevance weighting was not to remove stop words from the summaries before weighting. Consequently, a proportion of the suggested expansion terms consisted of labels such as, '*docno*' and '*bylin*'. This was shown to affect a small number of queries.

4.Experimental details

4.1 Subjects

All subjects were educated to graduate level in a non-computing, non-LIS discipline, and, with two exceptions, all our subjects were postgraduate students recruited from the Information Technology course at Glasgow University. None of the subjects had any formal training in information searching or retrieval, beyond basic training on the university library search facilities.

The average age of the subjects was 23 years, and the average previous search experience was 3 years. All subjects reported some experience with library systems but the majority of reported experience was gained using web search engines using a point-and-click interface. None of the subjects had used either the control system (ZPRISE), the experimental system or an IR system with summarisation facilities.

These subjects were relatively regular searchers, performing searches either daily or weekly, but were neutral as to how much they enjoyed the process of searching for information.

4.2 Technical drawbacks

There were two main technical problems observed in our experiments, both related to the length of time taken to produce summaries.

Summaries are produced on request, section 3.1, consequently speed is an important issue for this system. The average time taken to produce a summary was 5 seconds, the range being 0.5-20 seconds. However in practice summaries took longer than average to generate, somewhere of the order of 10 seconds on average. The main reason for this is that the summarisation time is dependent on the length of the original document.

In our experiments, the retrieved documents tended to be longer than average, consequently the summaries took longer than average to produce. This was criticised by several users, as discussed in section 5.2.

Secondly, summaries were requested by running the mouse pointer over a document title. This method was intended to be an intuitive method of obtaining summaries. However, as users were unfamiliar with the interface, they often scanned the pointer over several document titles unintentionally. This would not be a problem if summaries were created instantaneously but, as summaries took longer than expected to create, this often had the unfortunate side-effect of halting the interface until the summaries for the scanned documents had been produced. The user had then to wait until all summaries were created before being able to issue any more commands.

A further, although minor, problem was that as soon as the user moved the pointer away from a document title, the summary disappeared. This was criticised by several users who would have preferred the summary to remain visible until a new summary was requested.

5. Analysis

The experiments were run according to the matrix supplied by the track organisers, however technical problems meant that we were unable to run the complete set of subjects. Our final submission consisted of the full results for 10 out of 16 subjects[2]. Hence the following analysis is partial.

5.1 Quantitative analysis

5.1.1 Interface

To examine the novel features of our interface we extended the TREC post-system questionnaire to include specific questions on the use of summaries and RF.

The subjects were, on average, in favour of the summaries with an average score of 3.7[3] for the question "*Were the document summaries useful in answering the questions?"* and 3.5 for the question "*Were the document summaries a good representation of the full document text?"*. However, it is doubtful whether this second result was valid, as few users actually compared the full-text with the summary.

9 of 10 subjects judged the length of the summary as "*About right*", the remaining subject said the summaries were too short.

The subjects were less convinced about the benefits of RF with an average response of 2.7 for the question "*Was relevance feedback useful*?", 2.1 for the question "*Did the system add good terms to your query?"* and 2.5 for "*How well did you understand the relevance feedback option?"*.

RF was also shown to be unpopular in the exit questionnaires, we shall discuss this in more detail in section 5.2.

5.1.2 Topics

The subjects were generally unfamiliar with the topics before searching, the reported certainty before searching being on average between 1.24 – 1.9. The certainty of the users increased after searching for all topics. The final certainty ranged from 2.97 for topic 6, to 4.4 for topic 1. Although the pre-search certainty for both types of topics were roughly the same (1.60 for multiple part topics, 1.59 for comparison topics), searchers reported a greater degree of post-search certainty for multiple part topics (3.51 multiple part against 3.15 for comparison topics).

Although users found slightly more relevant documents for the multiple part topics (1.04 per search vs. 1.01 for comparison topics), the greater reported certainty seems to come from the interaction rather than search success. For the multiple part topics, subjects reported that these topics were easier to start a search on, and easier to search for[4]. However they reported that they were less satisfied with searches on the multiple part topics (3.00 vs. 3.27 comparison) and would have liked more time to search on these topics (3.41 multiple part vs. 3.45 comparison). There was a greater pre-search familiarity with the multiple part topics (1.96 vs. 1.78) not reflected in pre-search certainty.

5.1.3 Systems

Users marked slightly more relevant documents on average with the control system (1.04 vs. 1.01 per query), found it easier to start a search with the control system (3.58 vs. 3.36) and easier to search (3.05 vs. 3) but were overall less satisfied with the control system (2.62 vs. 2.51). The subjects felt they would have preferred more time with experimental system (2.62 vs. 2.81), possibly due to the time delay in producing the summaries.

From the exit questionnaires, the subjects claimed a relatively high level of understanding of the task (4.4), a fair similarity with other searching tasks (3.5). The subjects did not feel there was a great difference between systems (3.4).

---

[2] Five subjects started searching on the control system, five on the experimental system.

[3] These averages, and others in the remainder of the paper, are out of a possible 5, taken from the answers given in the standard TREC questionnaires.

[4] Ease to start (multiple part 4.46 vs. 3.97 comparison), ease to search (3.77 multiple part vs. 3.54 comparison).

Of the ten subjects tested 6 claimed the experimental system was easier to learn to use (1 for ZPRISE, 3 undecided), 7 found the experimental system easier to use (3 ZPRISE, none undecided), and 7 preferred the simplicity of the experimental system (3 opting for ZPRISE).

5.1.4 Search statistics

In this section we analyse the search statistics regarding the number of documents retrieved, query terms entered and documents assessed relevant for both systems. Table 1 summarises the basic search statistics.

	Control System	Experimental System
Initial query terms	3.54	4.69
Query terms added	1.24	1.55
Iterations (including initial ranking)	1.5	2.16
Documents assessed relevant by subject	1.04	1.01
Full texts viewed	2.94	1.19
Summaries generated	-	5.48

Table 1: Average search statistics per query

The searchers tended to use more query terms on the experimental interface than the control system and more terms were added through query expansion. Very few terms were added through the interactive query expansion facility. As noted before the searchers assessed slightly more relevant documents with the control system than the experimental system.

The searchers appeared to do more iterations of feedback with the experimental system although this is slightly deceptive as in fact they tended to do a new search more often than modify their existing query.

5.1.5 Search results

In Table 2 we outline the overall search results for our experiment. The users returned a higher proportion of non-supporting documents on the control system (41.67% - sum of columns 4, 7 and 8 in Table 2) than on the experimental system (36.84%), and a lower proportion of supporting documents with the control system (58.33% control vs 63.16% experimental – sum of columns 2 and 5 in Table 2).

	2:2	2:1	2:0	1:2	1:1	1:0	0:0
Control responses	8	-	-	6	-	1	9
Control average	33.33%	-	-	25.00%	-	4.17%	37.50%
Experimental responses	6	-	1	6	-	2	4
Experimental average	31.58%	-	5.26%	31.58%	-	10.53%	21.05%

Table 2: Summarised results – control system versus experimental system

		2:2	2:1	2:0	1:2	1:1	1:0	0:0
Topics 1 -4	**TREC average**	13.91%	2.16%	0.96%	28.06%	9.35%	9.83%	35.73%
	Our average	4.00%	-	-	48.00%	-	12.00%	36.00%
Topics 5 - 8	**TREC average**	46.60%	-	14.56%	-	-	-	38.84%
	Our average	72.22%	-	5.56%	-	-	-	22.22%

Table 3: Summarised results – Glasgow results versus average results

In Table 3 we compare our results from both systems with the average from the TREC participants. For the multiple part topics (1-4), our subjects returned a smaller percentage of fully supporting documents than average but a higher percentage of documents that supported a partial answer. The subjects also returned slightly more than average non-supporting documents.

For the comparison topics (topics 5 – 8) the subjects returned noticeably higher percentage of supporting documents and fewer non-supporting documents, suggesting that these topics were easier for our subjects, although in both our control and experimental systems our searchers returned more documents for the multiple part topics.

Topic	Control unique	Control total	Experimental unique	Experimental total	Total unique for query	Total TREC	Averages
1	1	5	1	3	1	13	7.62%
2	1	1	-	-	1	7	14.29%
3	2	6	2	5	3	17	17.65%
4	2	3	9	11	11	39	28.21%
5	3	7	3	4	5	7	71.43%
6	2	2	2	5	2	3	66.67%
7	5	9	1	1	6	23	26.09%
8	1	1	-	-	1	15	6.67%
Total	17	34	17	29	30		
Average	**50.00%**		**58.62%**				

Table 4: Document analysis results

In Table 4, we present an analysis of the overlap of the supporting documents found with the two systems. These results attempt to capture the performance of subjects using either system to discover new documents that support the answer to each query. Any document returned by a subject for a specific query that was marked by TREC assessors as 'supporting the right answer for this query' was used in the calculations, irrespective of the score assigned to that response. That means that even if a subject provided a wrong answer for a query based on a document marked as 'supporting' by the assessors, that subject would still be credited with finding a document that supports the correct answer for that query, and would be included in the calculations.

The **Control unique** and **Experimental unique** columns indicate the number of unique supporting documents found by subjects for each of the two systems (Control or Experimental) for a specific query. The **Control total** and **Experimental total** columns indicate the total number of documents marked by subjects for a specific query for each of the two systems. The **Total unique for query** column displays the total number of unique supporting documents for each query returned by both systems. Finally, the **Total TREC** column indicates the number of documents for each query that were marked as 'supporting the right answer' from the TREC assessors.

The results indicate that, on average, subjects using the experimental system performed better at discovering documents that could potentially support the right answer for a query. Subjects under both systems discovered the same number of unique supporting documents (17), however subjects using the experimental system discovered these documents in fewer attempts (29 vs. 34).

5.1.6 Search narrative

A question searchers have to answer is: *Which Children's TV program was on the air longer: the original Mickey Mouse club or the original Howdy Doody Show*?

The searcher starts with the query containing terms, *children's*, *TV*, *programs*, *Mickey*, *Mouse*, *Club*, *Howdy*, *Doody*, *Show*. That is basically used all the terms in the question description itself, suggesting the subject relied heavily on the task given. The system returned twenty documents titles and the searcher went through the list of titles in order.

The searcher viewed summaries of eighteen documents, these were not selected in the same order as the retrieved list. In addition, the searcher viewed the full text of three documents out of the twenty displayed.

There was no relevance feedback iteration and no query modification. The users returned the answer to the query as the Mickey Mouse show ran longer, based on the document, LA050690-0059. The confidence of the user in the answer is Neutral.

5.2 Qualitative

Relevance feedback was not popular amongst our searchers. As noted in the previous section searchers on the experimental system tended to enter new search terms rather than work with the modified query. We believe that the high familiarity with web search engines in our subjects may have promoted this behaviour. The subjects also claimed a poor understanding of relevance feedback. Each subject used the interactive option at least once. The RF option was also relatively unpopular on the control system.

Topic 1, "*What are the names of three US national parks where one can find redwoods*?" caused our subjects some difficulties as eight of the ten subjects returned document SJMN91-06312178 (shown in Figure 2), which discusses State rather than National Parks. The text of this document does make the distinction between the two categories of park, but we doubt whether the subjects read this document closely enough to pick up on this.

In answering the comparison topics, such as "*Which painting did Edvard Munch complete first: "Vampire" or "Puberty"?*", several subjects did not find any individual document that supplied an answer but managed to find an answer by analysing information found in more than one document. For example, on the Munch topic, several users found a date for the completion of "*Vampire*" in one document, a date for "*Puberty*" in a different document and answered the question citing both documents as evidence.

Although our analysis is not conclusive, we believe that summaries may have resulted in some false positive answers. In this case, the searcher assesses the summary relevant without reading the full text of the (non-supporting) document.

A good example of this is the assessment of document AP890215-0071 for the topic "*Name four films in which Orson Welles appeared*". One summary produced for this document contained the sentence "*Turner Entertainment Co. said it will not colorize Orson Welles' black-and-white film classic "Citizen Kane" because the late director's estate may have the right to prohibit it.*" which correctly identified the film Citizen Kane as being one of the films of Orson Welles.

However the summary also contained the last sentence of the document, "*Movie purists have previously lamented the colorizing of such classics as "It's a Wonderful Life", "Casablanca" and "A Christmas Carol.*"" which led the searcher to credit Mr Welles as appearing in these films, even though this was not supported by the text of the full document. Although this may have affected some of our searchers, it may not be a real concern in an operational environment in which users are not so restricted by time limitations.

All subjects liked the use of summaries but felt the summaries took too long to produce. The subjects also liked the simplicity of the interface event if they ignored the RF and query expansion options.

6. Conclusion

Our research aim was to investigate the use of summarisation techniques for a question-answering task. Although our subjects returned a low number of documents, the analysis showed that subjects returned a higher proportion of supporting documents and a lower proportion of non-supporting documents with our summarisation system. The subjects also viewed fewer full documents per relevant document found with the experimental system than the control system. Although our experiments were only partially completed, these two findings indicate that our experimental hypothesis that summaries can help users target relevant documents and eliminate non-relevant documents more effectively is worth investigating further. In addition, the positive response from our subjects towards the use of summaries indicate the summaries are not only effective but can also be a popular aid to searching.

Acknowledgements

Ian Ruthven is currently supported by the Library and Information Commission funded project *'Retrieval through explanation'*, Anastasios Tombros is supported by a University of Glasgow Postgraduate Scholarship. We would also like to acknowledge the Computing Science Research Committee for funding this research and all the experimental subjects who took part in these experiments.

References

[BMR95] R. Brandow, K. Mitze, L. and Rau. Automatic condensation of electronic publications by sentence selection. *Information Processing and Management*. **31**. 5. pp 675 - 685. 1995.

[Ed64] H. Edmundson. *Problems in automatic abstracting*. Communications of the ACM. 7. **4**. pp 259 - 263. 1964.

[Efth95] E. N. Efthimiadis. *User-choices: a new yardstick for the evaluation of ranking algorithms for interactive query expansion*. Information Processing and Management. 31. 4. pp 605 - 620. 1995.

[MHKHOFCS98] I. Mani, D. House, G. Klein, L.Hirschman, L. Obrst, T. Firmin, M. Chrzanowski, and B. Sundheim. *The TIPSTER SUMMAC text summarization evaluation: final report*. MITRE Corporation Technical Report. 1998.

[PG88] M. Porter and V. Galpin. *Relevance feedback in a public access catalogue for a research library: Muscat at the Scott Polar Research Institute*. Program. **22**. 1. pp 1 - 20. 1988.

[RSZ71] J. Rush, J, R. Salvador and A. Zamora *Automatic abstracting and indexing. II. Production of indicative abstracts by application of contextual inference and syntactic coherence criteria*. Journal of the American Society for Information Science. 22. **4**. pp 260 - 274. 1971.

[Tom97] A. Tombros. *Reflecting User Information Needs Through Query Biased Summaries*. MSc Thesis, Technical Report (TR-1997-35) of the Department of Computing Science at the University of Glasgow, Glasgow G12 8QQ, UK. 1997.

[TS98] A. Tombros and M. Sanderson. *The advantages of query-biased summaries in Information Retrieval*. Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 2 - 10. 1998.