A Comparative Study of Online News Retrieval and Presentation Strategies

Steven Morrison, Joemon Jose
Department of Computing Science
University of Glasgow
Lilybank Gardens
Glasgow
G12 8QQ
(morrison, jj)@dcs.gla.ac.uk

Abstract

We introduce a news retrieval system on which we evaluated three alternative presentation strategies for online news retrieval. We used a user-oriented and task-oriented evaluation framework. The interfaces studied were Image, giving a grid of thumbnails for each story together with query-based summaries presented as tooltips, Summary, which displayed the summary information alongside each thumbnail, and Cluster, which grouped similar stories together and used the same display format as Image. The evaluation showed that the Summary Interface was preferred to the Image Interface, and that the Cluster Interface was helpful to users with a set task to complete. The implications of this study are also discussed in this paper.

1. Introduction

With the rapid growth of technology, both in power of computers and the underlying infrastructures, large archives of video information are created. News broadcast is one of the most prominent types of digital archive. Such archives have become common and have contributed to information overload [8]. The advent of 24 hour TV news channels has dramatically increased the available information and in doing so has placed increasing demands on viewers to filter out irrelevant stories and make optimal use of their time.

There are many situations where the nature in which video is transmitted, with no easy means of pausing or seeking through the video present, is not sufficient to meet a user's needs. An example of such an occasion is a user who is researching a current event, but is unable to record the facts presented at the same pace as they are being streamed. TV based news offers only sequential access to news, forcing users to wait for the stories they are interested in.

Information retrieval tools can be applied to deal with this problem. However, the specific nature of video information demands novel processing and presentation strategies. Previous work on the NewsFlash system yielded a suitable online news system for storing the news footage, but did not investigate the format of search results displayed to the user [5]. Work on the Fischlár system created a novel format for news video retrieval, but the representation of search results may have been too condensed for users to make good use of [7]. Similarly, the ANSES system provides an innovative method for generating news summaries based on a query, but no studies have been carried out regarding the presentation of these summaries to the user [9].

With these matters in mind, an evaluation was carried out on three interfaces to a news search engine, VideoSqueak, to ascertain the respective merits of each interface in an attempt to develop a user interface suitable for users of varying experience with such systems.

In this paper, the system under investigation, VideoSqueak, will be described, together with the three interfaces created for the evaluation. This evaluation will then be explained, the results presented and the conclusions and implications of this study given at the end of this paper.

2. The VideoSqueak System

We have developed a web-based news search engine, VideoSqueak, which is based on [5]. VideoSqueak records continuous news broadcasts and facilitates online querying and retrieval of the video archive created. As with many such systems [5] [6] [7] [9], VideoSqueak can be best described in two separate parts, namely the archive preparation and the web interface.

VideoSqueak records the evening news from BBC1, including both the national news and the local (Scottish) news that follows, totaling an hour of news per day. Coinciding with this is the recording of the subtitles that accompany the news to give the textual content of the news.

Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering (ISMSE'04)
0-7695-2217-3/04 $20.00 © 2004 **IEEE**

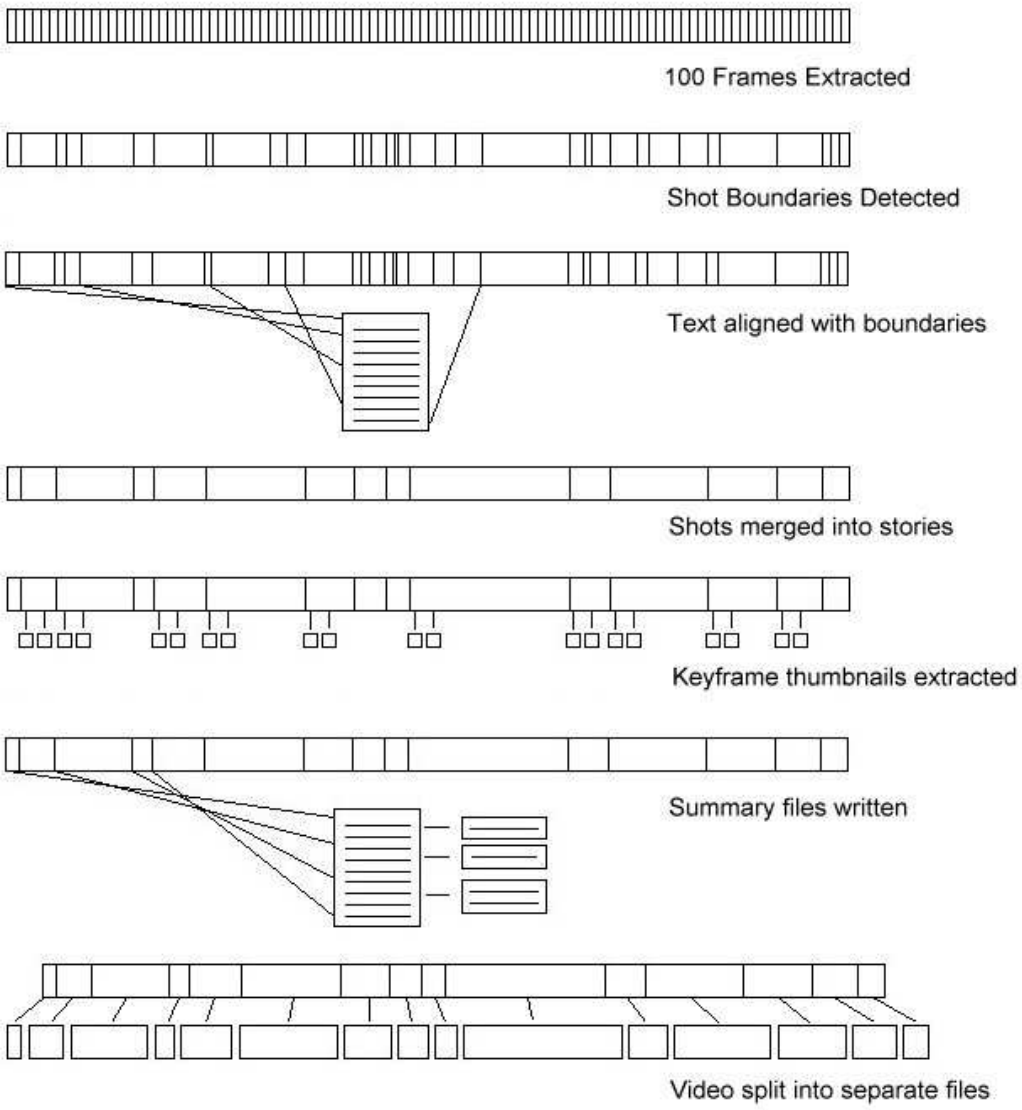

Figure 1. Archive Preparation Overview

Initially, the video file is split into shots. Shot detection is the process of detecting shot boundaries in a video file, and has been the subject of previous research projects [4]. A shot is a single continuous collection of frames from a single camera feed, and may contain camera movements including panning and zooming. Though many forms of shot detection exist that can improve precision and recall values for a search [3], previous work has shown that colour histogram comparisons are adequate for most systems, and can be performed faster than most other methods [11]. This method was therefore chosen for the VideoSqueak system.

The cleaned subtitle text is then aligned with the detected shots before the shots are merged into stories, using the assumption that two adjacent shots are from the same story if they contain ten common words. This process gave story detection reliability that was sufficiently accurate for our needs. Two thumbnails per story are extracted from the original video file to represent the stories in the web interface. Each new story is then added to the archive for later querying and retrieval.

The entire archive preparation process is summarised in Figure 1. With these stories added to the system's archive, the issue of locating relevant stories will now be discussed.

2.1. Web Interface

The idea of our system is to make news access online and hence a web interface has been created. When creating a web-based search engine, a designer must be aware of the volume of data that they use to illustrate results. In the case of a video search engine, using video footage would be difficult due to the size of video files required being too cumbersome for dial-up connections to handle. Therefore other information must be used, but deciding on the right information to present, and the format in which to present it, can be difficult tasks [6].

Presenting the user with images extracted from a story's video can allow them to quickly deduce key elements contained in the story, and so assign an initial value of relevance to the story. Further information on the story's content can then give the user more insight into its content, and permits them to make an informed decision regarding whether they wish to view the story in video form. In VideoSqueak, this extra information is given in the form of a summary for each story, generated with the sentences that include the given query terms most often.

With the content to use in search results decided, three interfaces were created for the VideoSqueak system, and these are described below.

2.1.1. Image Interface The results layout offered a series of images, one per story, as a 4*3 grid, to allow the results to fit into the space allocated to results without the need for scroll bars. Moving the mouse pointer over an image caused the image to be replaced by a second image for that story, and for a text tooltip to be displayed. The text in the tooltip was comprised of the 4 sentences found to be most relevant to the user's query. Navigation links were added to the results display area when more than 12 results were returned for a query, to allow a user to move between successive results pages easily. The result display style for the Image Interface is shown in Figure 2, which gives an overview of the whole VideoSqueak interface. This interface style is similar to that used in [5].

2.1.2. Summary Interface The motivation behind this interface was the belief that the Image Interface described above represented each result in a way that was overly compact, with each textual summary hidden from the user until the mouse was hovered over a result's image. Also, the use of tooltips to show summary text was felt to be detrimental to the system, as the use of tooltips to display the content of an image's ALT tag was not a feature given in many popular browsers. Even though MS Internet Explorer did support this feature, a tooltip would be displayed for only a few seconds. Given the length of the summaries generated by VideoSqueak, this brief period was felt to be too short for a user to read the entire content of the tooltip.

To combat this problem, the use of tooltips was removed from the Summary Interface. Each result was instead displayed as a single entry in a table's row, and represented by two images as before, with the textual summary shown to the right of these images. Also, the terms used by the user to perform their search were highlighted in this text, in order to aid the reading of the text and to illustrate why each story in the result set had been chosen for inclusion in the results.

Due to the increased amount of space each result required in this interface when compared to the Image Interface, the number of results per page was reduced to 10. An example of this display style is shown in Figure 3.

2.1.3. Cluster Interface The third interface created for VideoSqueak was an interface designed to display search results in the form of clusters. A cluster is a group of results with similar content. Clustering is an area of research that has been heavily studied in the field of information retrieval, and while many algorithms for clustering documents exist [12], few popular free-text search engines use clustering.

For this system, the similarity of two stories was based on their textual content. We used the WPQ algorithm to select the most important terms for each story, then used these to determine the similarity of two stories [10]. The clusters were generated when a search was performed, causing a delay of a few seconds more than the previous two interfaces.

This Cluster Interface is illustrated in Figure 4. Each story was represented in the same way as in the Image Interface, with two images and a tooltip giving the textual summary. Each cluster was represented as a row of results in this form, with a maximum of four stories per cluster. Each cluster was also labeled with a set of terms found to be common to all of the contained stories.

By grouping similar stories together within a set of search results, the intention was that users would be able to identify in a shorter time the stories that were relevant to what they were interested in, than if they were to use a standard list of stories, ranked independently by their relevance to the users query.

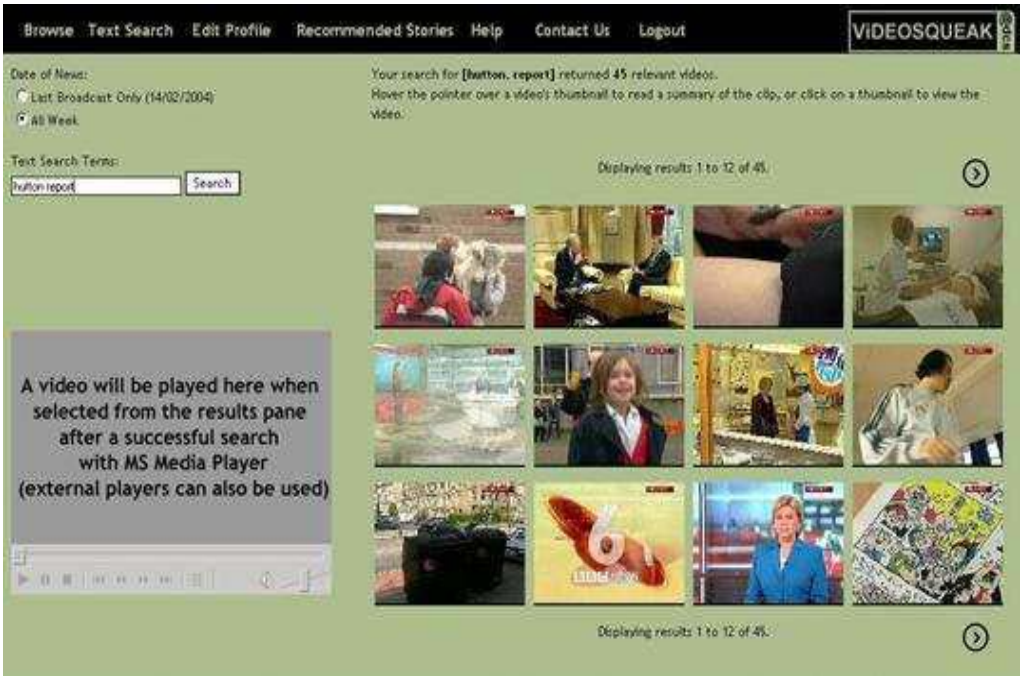

Figure 2. Image Interface - Result Display Example

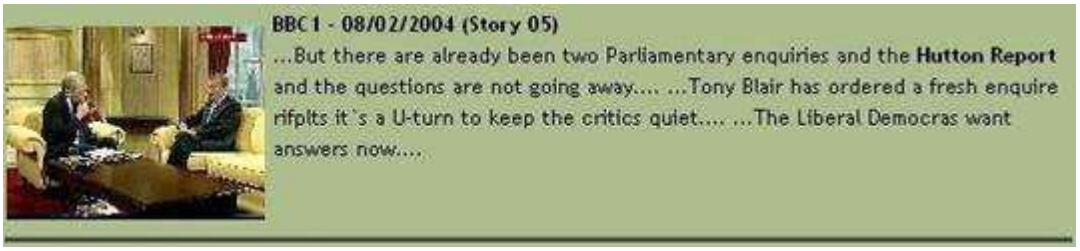

Figure 3. Summary Interface - Result Display Example

Figure 4. Cluster Interface - Result Display Example

2.2. Adaptive Search

The VideoSqueak system included an adaptive search that was performed after each story selection by the user. This search would be based on the search terms originally given by the user, as well as the most commonly occurring words from the chosen story, in an attempt to find relevant stories that were similar to the story chosen for viewing.
This feature was used for both the Image and Summary interfaces, but was not offered in the Cluster interface. The decision to remove this feature from the Cluster interface was based on the fact that the videos similar to the story chosen would already be shown within that video's cluster, and so a further search for these videos would not be necessary.

3. Evaluation

In order to evaluate the suitability of these interfaces for online searching, we conducted a user evaluation. We followed a task-oriented and user-oriented evaluation framework [1] [13].
From the above descriptions of the interfaces and issues under consideration, we derived the following hypotheses:

- The Summary Interface imporoves the search experience over the Image Interface, by presenting the results in a more readable form.
- The Cluster interface is helpful to users, and by extension improves their searching experience.

3.1. Experiment Methodology

The evaluation involved recruiting a set of volunteers who were potential users of the VideoSqueak system, were it made publicly accessible. These participants were initially asked a set of questions relating to their interest in current events, and their past experience of search engines. Each participant was then asked to use each of the three interfaces described in Section 2.2 to complete a different task, after being introduced to this interface and shown how to interact with it. Each task description included a context, such as asking the user to assume they were a writer for the University's newspaper, to help the user understand what they were required to do in each task.
Every user was asked to use each interface within a single session. To avoid familiarity with the contents of the archive affecting a user's ability to find relevant stories for a task, three tasks were devised for use in the evaluation. These tasks applied to different areas of current events (namely crime, politics and sport). By choosing such diverse tasks, a user's background knowledge could be assumed to not cover each area fully, while also ensuring that no story within the archive would apply to two different tasks.
After each task, the users were asked a set of questions to determine their impressions of the VideoSqueak system in general, as well as the interface they had just used. At the end of the three tasks, each user was asked to state which interface they found to be most helpful and which was least helpful when completing the task set, and to qualify their choices with reasons. The same style of questioning was used to find the interface each user liked the most, and which they like least. This form of evaluation, with the different interfaces being used by the same users for different tasks, allowed for repeated measures analysis to be carried out on the results, with the interface type used as an independent variable.
When performing user-based evaluations over a set of different interfaces of a system, a risk is present that if all users are introduced to the interfaces in the same order, their familiarity with the system's characteristics will increase with each interface they use, and so their confidence in the system will grow, causing them to alter their interaction style. This could lead to the last interface to be introduced to the users being rated the most liked or helpful. A similar problem exists with the tasks given to carry out with the systems. In this case, a task that is found by the majority of users to be more difficult than the other two given could adversely colour the users' opinions of the interface they were given to complete the task with. Therefore, each task must be attempted against each interface an equal number of times.
To combat these potential problems, a Greco-Latin square of the interfaces, tasks and the order of presentation to users was built. A standard Greco-Latin square is a matrix where every Latin character appears once in each row and once in each column, the same is true for each Greek character, and the condition that no Greco-Latin character combination occurs more than once holds. When applied to the above problem, the Greco-Latin square becomes a cubic matrix, with an index for each of task, interface and position in the order of introduction to user.
The entire evaluation process described above was carried out on a volunteer as a pilot run to ensure that the tasks could be completed and were not too complex. A group of 18 users was then recruited to participate in the evaluation, by emailing students within our department. Each participant was given a written explanation of the content of the experiment, which they signed, before the experiment commenced, to ensure compliance with the conditions of the experiment.

3.2. Results

With the evaluation complete, formal and informal analysis of the results obtained was performed. The findings of this analysis will now be discussed.

3.2.1. Comparison of Systems When asked which system was found to be the most helpful to the users, 11 of the 18 participants answered the Summary Interface, while only one said that this was the least helpful system. The Image Interface was found to be the least helpful by 11 people, and only one person felt it was the most helpful. The Cluster Interface was found to be the most helpful by 6 users, and least helpful by 7. These values give weight to the argument that the Summary Interface is the most helpful of the three, while the Cluster interface was still found to be helpful by many, and in general was favoured to the Image Interface.

The second question in this part of the evaluation asked users to state which system they had liked the most out of the three given, and which system they had liked the least. The results obtained here were very similar to those for the previous questions regarding helpfulness, with the exception of the number of participants who were most favourable to the Cluster Interface, which fell from 6 to 3, with two of the votes transferred to the Summary Interface and one to the Image Interface. This change in opinion between the two questions for some users would suggest that although the Cluster Interface was found to be helpful, users preferred the results being displayed in the more traditional style of ordering by relevance to the query.

For each question, the participants were asked to elaborate on the systems they chose to give as answers. In these explanations, the most frequent complaint about the Image Interface was how difficult it was to read the story summaries, while 13 users cited improved readability as a benefit of the Summary Interface. According to these users' comments, readability was aided most by the summaries being permanently available on screen, and the highlighting of keywords. The layout of the results was another area in which the Summary Interface was felt to be superior to the Image Interface.

The reasons given for and against the Cluster Interface were more varied than for the other two. Eight users found the clustering of videos to be advantageous, while others were against this form of display. Three users found the clustering process to be confusing, though the removal of the reordering of results from this interface was found to be appreciated by some users, who also expressed a dislike for this behaviour in the Summary and Image interfaces. One user suggested the inclusion of "cluster summaries" to give a better overview of each cluster's content.

3.2.2. Post Search Questionnaire At the end of each task, the participants in the evaluation were asked to give comments on the system they had just used. We used Likert scales, semantic differentials and open-ended questions in these questionnaires. Of the 18 people recruited to take part in the evaluation, 8 said that they were either confused or annoyed by the automatic reordering of search results that was used in the Summary and Image interfaces when a story was selected for viewing as a video clip. Many of these people suggested that the user could be given the choice of whether this feature was in use, or that the user could explicitly request the reordering to occur if they wished to use it following a story selection. Seven of the participants stated that using tooltips to display story summaries was detrimental to the Image and Cluster interfaces. The main reasons given for this were that the summaries were not permanently displayed on the screen, the tooltip text was not easy to read, and the tooltips would disappear after a few seconds of viewing.

Using the Friedman test, only the question regarding how clear the results formatting was for each system was found to give reasonably conclusive results. In this case, it was shown that the Summary Interface was preferred over the Cluster interface by the majority of users. All other questions yielded values that indicated the results were too similar to rule out the possibility that they were caused by chance.

The fact that only one question gave a strong indication of a common opinion among the majority of users could be due to one of many reasons, which will be discussed in Section 5.

To summarise, the Summary Interface was found to be more appropriate for use in this context, while the Cluster Interface proved to be useful and appreciated by some users.

4. Discussion and Conclusions

The Summary Interface was preferred to the Image Interface by the vast majority of participants in the evaluation. The reasons for this overall opinion given by the participants suggest that in order to improve the Image Interface, the tooltip summaries should remain on screen for as long as the user required, instead of disappearing after a few seconds. The terms from the user's search could also be highlighted in the tooltip as they were in the Summary Interface. The Summary Interface was also preferred to the Cluster Interface, but many users found the clustering of stories to be helpful. Users felt that future development of this interface would be improved with an explanation of the clustering process, and more textual content for each cluster available for display. Therefore the second hypothesis of the evaluation was found to not hold.

The Post Search Questionnaire given to each participant after each task contained 26 questions that used Likert scales or semantic differentials to label responses, as well as an area in which participants could leave comments on the system last used. It could be argued that this many questions made the evaluation process too cumbersome for the participants. Each question in the Post Search Questionnaire, aside from the final open-ended question, used a scale with 10 possible values to choose from (0 to 9 inclusive). An even number of values was chosen to ensure that no middle ground existed for any question. This forced the participants to answer each question in a positive or negative manner, with no neutral territory. In a situation where a user wished to answer neutrally, they were still required to bias their answer towards the positive or negative end of the scale given. Despite this conscious effort to improve the results obtained from the questionnaires, it was later found that Likert scales with 5 or 7 points have been proven to be the most effective for measuring users' opinions [2]. While it is unlikely that reducing the number of available values in each question would have altered the overall results obtained, the neutral territory option for each question may have been a popular choice among participants, and so including this option is perhaps beneficial in this type of evaluation.

5. Future Work

The role of adaptation in the VideoSqueak system shall be researched further with the introduction of adaptive user profiles for the recommendation of new stories to users based on their preferences. These recommendations shall be made via email or using the Standard Messaging Service for mobile phones, depending on the user's preference. This extension will give rise to the development of many new interfaces, with the aim of making VideoSqueak available from multiple devices and settings. The system shall also be extended to include news recorded throughout the day, and an Interactive TV interface as an alternative interaction format.

6. Acknowledgments

The authors would like to thank Alan Haggerty for the permission to build upon his original system, and Marcus Pickering for his assistance in teletext capture.

References

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

[2] J. D. Browne and M. Swan. *Understanding research in second language learning: A teacher's guide to statistics and research design*. London: Cambridge University Press, 1988.

[3] P. Browne, A. Smeaton, N. Murphy, N. O'Connor, S. Marlow, and C. Berrut. Evaluating and combining digital video shot boundary detection algorithms, 1999.

[4] J. Calic and E. Izquierdo. Towards real-time shot detection in the mpeg-compressed domain. *Proceedings of WAIMIS*, pages 1390–1399, 2001.

[5] A. Haggerty, R. W. White, and J. M. Jose:. Newsflash - tv news delivery on the web. In *1st International Workshop on Adaptive Multimedia Retrieval*, pages 72–86, 2003.

[6] Ide, H. Mo, N. Katayama, and S. Satoh. Topic-based inter-video structuring of a large-scale news video corpus. *Proc. 2003 IEEE Intl. Conf. on Multimedia and Expo (ICME2003), vol.3*, pages 305–308, 2003.

[7] H. Lee, A. Smeaton, C. O'Toole, N. Murphy, S. Marlow, and N. E. O'Connor. The fi schlár digital video recording, analysis and browsing system. *Proceedings of RIAO2000: Content-Based Multimedia Information Access*, pages 1390–1399, 2000.

[8] M. R. Nelson. *We Have The Information You Want, But Getting It Will Cost You: Being Held Hostage By Information Overload*. Crossroads, 1994.

[9] M. J. Pickering, L. Wong, and S. M. Riger. Anses - summarisation of news video. In *International Conference on Image and Video Retrieval (CIVR-2003)*, pages 425–434, 2003.

[10] S. Robertson. On term selection for query expansion, 1990.

[11] A. F. Smeaton. Indexing, browsing and searching of digital video and digital audio information. *ESSIR 2000*, pages 93–110, 2000.

[12] P. C. J. van Risjbergen. *Information Retrieval*. Cambridge University Press, 1999.

[13] R. W. White, I. Ruthven, and J. M. Jose. The use of implicit evidence for relevance feedback in web retrieval. *Proceedings of the 24th BCS-IRSG European Colloquium on IR Research*, pages 93–109, 2002.