Multi-Document Summarization: Methodologies and Evaluations

Gees C. Stein, Amit Bagga and G. Bowden Wise

General Electric, Corporate R&D, One Research Circle, Niskayuna NY 12309, USA
{steing, bagga, wiseg}@crd.ge.com

Abstract

This paper describes a system for the summarization of multiple documents. The system produces multi-document summaries using clustering techniques to identify common themes across the set of documents. For each theme, the system identifies representative passages that are included in the final summary. We also describe a methodology for evaluation of our system which is based upon a question answering task. Results of our evaluation are also presented.

1. Introduction

Multi-document summarization differs greatly from single-document summarization. In fact, single-document summarization can be considered as one of the critical sub-tasks of multi-document summarization. However, there exist several other important sub-tasks including identification of important common themes or aspects across the documents, selecting representative summaries for each of these themes, and organizing the representative summaries for the final summary. In addition to these sub-tasks, an important aspect of building a multi-document summarization system is its evaluation. In this paper we describe a system for multi-document summarization. We also describe a methodology for the evaluation of our system and present results using this evaluation methodology. We will conclude this paper with related work and future plans.

2. Single-document summarization

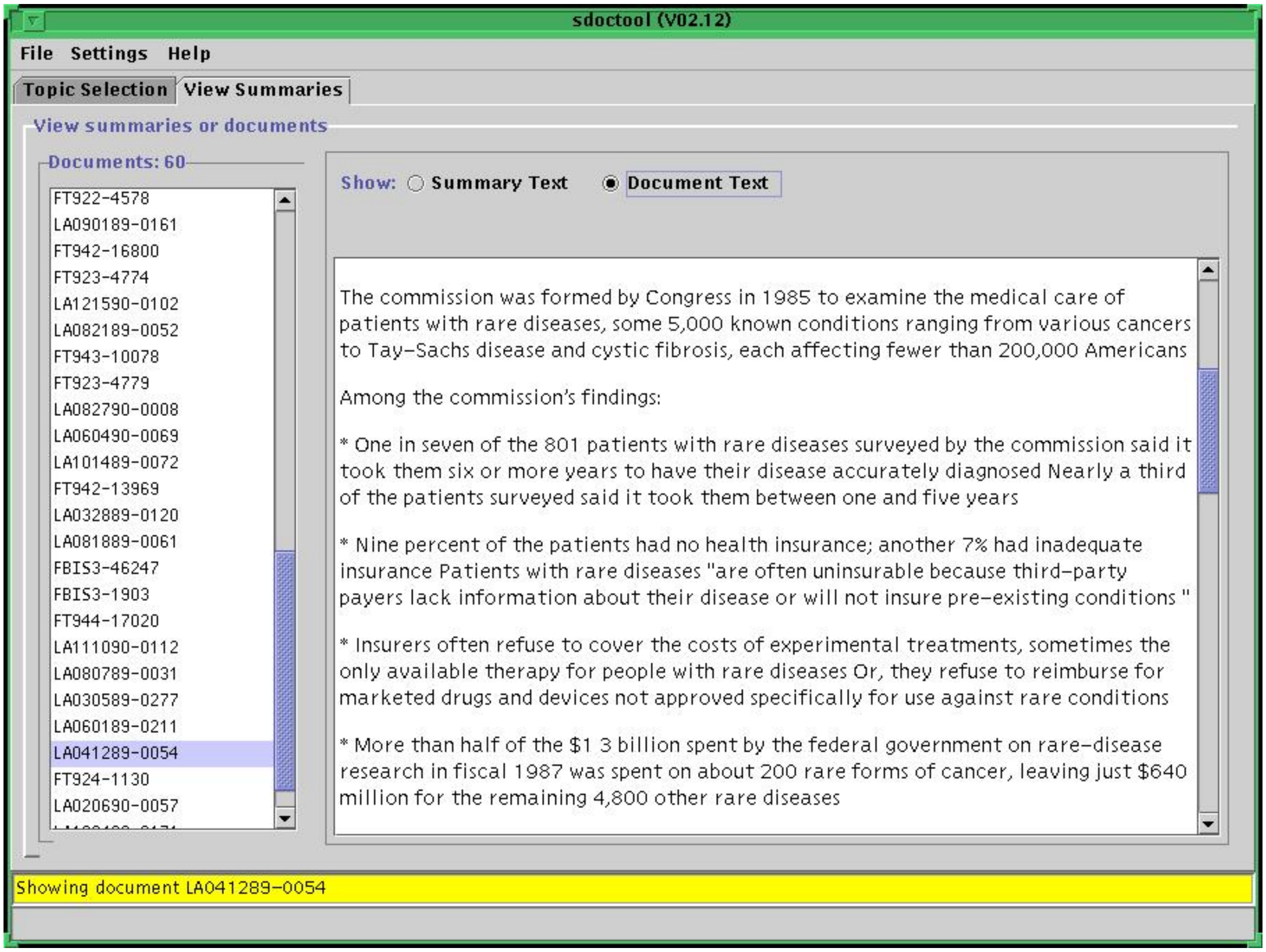
2.1 The Approach

Our multi-document summarization system uses the output of our single-document summarizer (Strzalkowski et al.1999 ; Strzalkowski, Stein and Wise 1998, Strzalkowski, Wang and Wise 1998). The system takes as input the document to be summarized, a target percentage for the desired summary length, and a topic description. First, the system segments the text into text-units, in our case, paragraphs. Next, the system reconnects passages that contain backward links such as anaphors. Reconnected passages will always appear together in the final summary, if selected. In other words, they cannot be separated from each other. The topic description is used to form a set of query terms which are used, together with the target length, to score each passage. The system iteratively scores combinations of passages, stopping when no improvement in score is reached anymore. The set of passages with the best score is the final summary. The system does not depend on any statistical data taken from a similar-text corpus.

Most other text-extraction based approaches select key sentences to build summaries (e.g., Luhn 1958, Paice 1990, Rau, Brandow & Mitze 1994; Kupiec, Pedersen & Chen 1995). The drawback of this approach is that the resulting summaries often lack coherence and can be hard to read. The paragraph-based approach produces summaries that are better readable and more coherent since a paragraph, in general, is a self-contained unit and its relationships to surrounding paragraphs are easier to trace.

2.2. The Single-DocumentTool

Initially, the user selects a document set and provides a topic description. This brings up a screen with two main windows (Example 1). The left window lists all documents in the set; clicking on any of them will bring up the summary in the right text window. Using the radio buttons on top of this window the user can either view the full-text version (the source) or the summary of this document.



FT922-4578
LA090189-0161
FT942-16800
FT923-4774
LA121590-0102
LA082189-0052
FT943-10078
FT923-4779
LA082790-0008
LA060490-0069
LA101489-0072
FT942-13969
LA032889-0120
LA081889-0061
FBIS3-46247
FBIS3-1903
FT944-17020
LA111090-0112
LA080789-0031
LA030589-0277
LA060189-0211
LA041289-0054
FT924-1130
LA020690-0057

The commission was formed by Congress in 1985 to examine the medical care of patients with rare diseases, some 5,000 known conditions ranging from various cancers to Tay-Sachs disease and cystic fibrosis, each affecting fewer than 200,000 Americans

Among the commission's findings:

* One in seven of the 801 patients with rare diseases surveyed by the commission said it took them six or more years to have their disease accurately diagnosed Nearly a third of the patients surveyed said it took them between one and five years

* Nine percent of the patients had no health insurance; another 7% had inadequate insurance Patients with rare diseases "are often uninsurable because third-party payers lack information about their disease or will not insure pre-existing conditions "

* Insurers often refuse to cover the costs of experimental treatments, sometimes the only available therapy for people with rare diseases Or, they refuse to reimburse for marketed drugs and devices not approved specifically for use against rare conditions

* More than half of the $1 3 billion spent by the federal government on rare-disease research in fiscal 1987 was spent on about 200 rare forms of cancer, leaving just $640 million for the remaining 4,800 other rare diseases

Showing document LA041289-0054

Example 1. The Single-Document Tool

3. Multi-Document Summarization

3.1 The Approach

The development of our multi-document summarizer is based on a few simple initial assumptions. The documents to be summarized are text-only, news documents that are well formatted. The goal is to create indicative summaries which give the users the gist of the original documents, i.e., if of interest, the user can decide to read particular full-text documents for more details. More on the system can also be found in Stein, Strzalkowski and Wise (1999 ; 2000(to appear)) and Stein et al. (2000)

Our basic approach attempts to generate a text summary while avoiding the repetition of identical or similar information and presenting the information in such a way that makes sense to the reader. With this in mind we decided on the following basic algorithm:

1. Summarize each document
2. Group the summaries in clusters
3. For each cluster select representative passage(s) that will contribute to the final summary
4. Organize these passages in a logical way.

3.1.1. Create individual summaries

The first step of the process of generating a multi-document summary is to create individual single- document summaries for all documents in the set.

This is done by creating a topical summary of 15% length using the user-specified topic and our SD (single-document) summarizer. Documents that seem to be irrelevant to the topic nevertheless result in a short default summary. In other words, no documents are filtered out using the SD summarizer. The 15% is chosen based on past evaluation results (Strzalkowski, Stein and Wise 1998).

3.1.2 Group Summaries

The second step of the multi-document (MD) summarization process is the grouping of the individual summaries into clusters. At this point, the system decides the SD summaries that are relevant to the topic, since only these are used for the final summary. The final summary contains only the main topics covered by the documents since repetition or very similar topics do not add much extra value to the summary. Therefore, documents are 'clustered' on the basis of the contents of their summaries where a cluster consists of summaries that describe a similar topic. For those documents that seem to discuss a similar topic, representative segments are eventually chosen for the MD summary while the other ones are 'hidden', i.e., not shown but still accessible to the user.

Since the final multi-document summary is highly dependent on the clusters, we experimented with a variety of approaches for producing the clusters. These approaches use the same basic similarity metric, Dice's coefficient (Van Rijsbergen, 1979), after stemming and removing stopwords, to compare two summaries *S1* and *S2*. If *sim(S1,S2)* is larger than a certain threshold the two summaries are considered to be similar.

We tried several graph-theoretic approaches that make use of a similarity matrix, where the nodes in a graph correspond to the individual summaries while an edge between two nodes corresponds to the similarity of the two corresponding summaries being above a certain user-defined threshold. It should be noted that the resulting graph may consist of several disjoint sub-graphs.

In our first approach, in order to find coherent clusters, we find *all* maximal complete subgraphs (cliques) in this graph (Van Rijsbergen 1979, Everitt 1993). We experimented with two different methods of postprocessing to merge those clusters that are very similar. This allows for overlapping clusters. The first method merges clusters that seem to address the same concept based on the words all members of a cluster have in common. If one cluster's common-word set is a subset of another cluster's common-word set, they are merged. The second method considers the number of members one cluster has in common with the other.

If this number is higher than a certain percentage (variable), they are merged. We chose 50% as the cut-off for merging clusters in this method. Experiments showed that the second method of merging clusters resulted in clusters of better quality. Documents relevant to the same topic were distributed over fewer clusters, and clusters were purer, i.e., less frequently would they contain documents from different topic sets (Stein et al., 2000).

As a second approach, we implemented a basic hierarchical complete-link algorithm (Salton 1989), which is more time-efficient than the clique-based approach. The algorithm starts with single-member clusters and merges clusters with the highest similarity until no clusters can be merged. In our case this happens when no two clusters have a similarity value above the user-defined threshold. Cluster similarity is equal to the *lowest* value of similarity between any pair of elements *e1* and *e2* from respective clusters. The resulting clusters are complete subgraphs that are non-overlapping.

It should be noted that both approaches are order-independent, i.e., it does not matter in what order the documents are processed. Earlier experiments (Stein et al., 2000) showed that for the same threshold the clique-based approach did better than the complete-link approach. The additional benefit of the clique-based algorithm is that it allows for overlapping clusters, which makes sense since documents can be relevant to more than one topic. However, the complete-link algorithm is considerably faster and creates similarly good clusters for lower thresholds. The user can choose between both approaches depending on the task and document set at hand. For the purpose of the experiments described later in the paper we used the complete-link approach.

3.1.3 Select representative passages

The third step of the multi-document summarization process involves selecting a member of a cluster as a representative summary for the cluster. When the user wants a topical summary, the topic description is used to pick the summary that has most similarity to the topic. In the case of a generic summary, the representative summary chosen is one that best represents the cluster. In this case, the summary that has most occurrences of the common terms across documents in a cluster is chosen. Since clusters can be overlapping it is possible that the same segment(s) is chosen to represent several clusters.

We are currently investigating selecting one or more passages as opposed to selecting an entire summary as the representative summary.

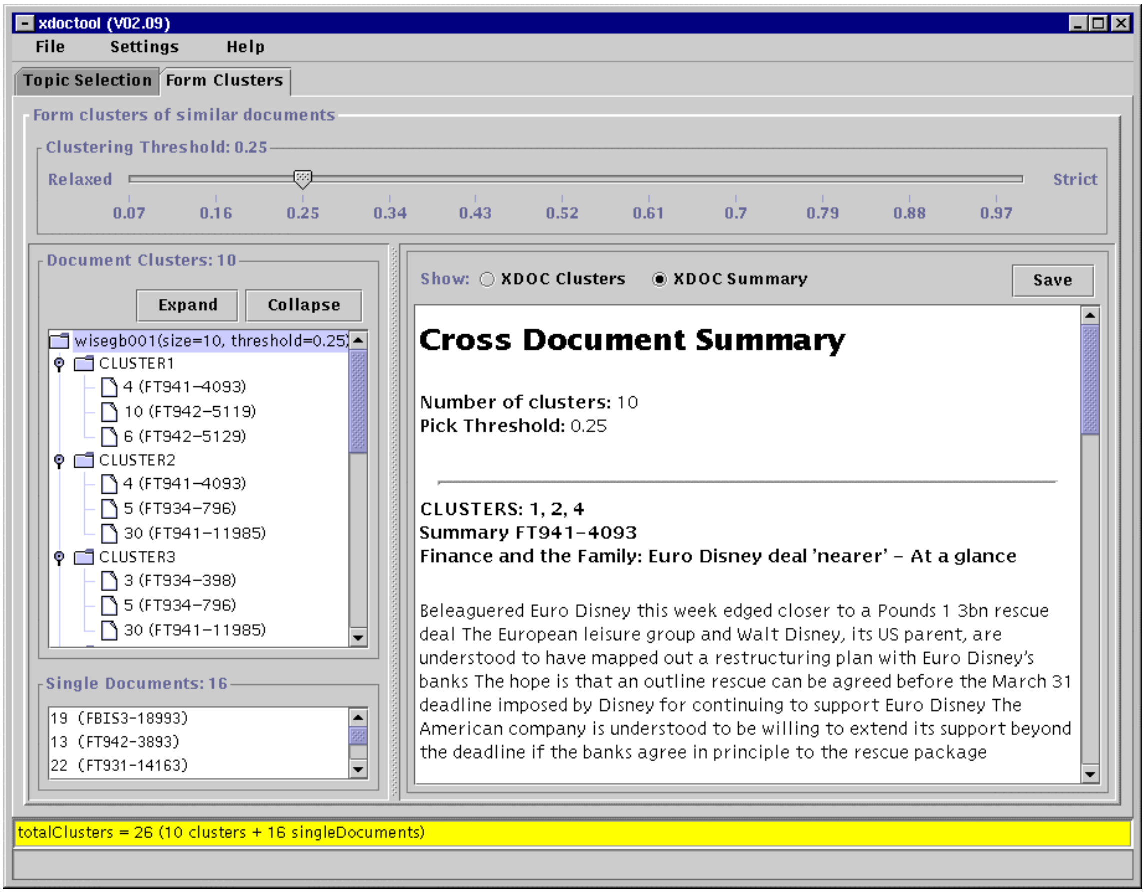
3.1.4 Organize selected passages

Finally, the last step of the multi-document summarization process involves organizing the selected passages in an order that makes sense to the reader. Such an order might depend on the topic (in case of topical summaries), the user and the task at hand, among others. Currently, we organize the selected passages based upon topic similarity. For generic summaries, the organization involves placing "similar" documents together so that all information about a particular topic is placed contiguously. This is done using the same similarity metric mentioned above to compare two documents: similar passages are placed close to each other. Single-document clusters are shown separate.

For topical summaries we use similarity to the query description to order the representative passages such that the most relevant passages are placed on top.

3.2. The Multi-Document Tool

The system, named XDOCTOOL, consists of two main screens. The first lets the user select a collection of documents (by pointing at a directory containing those documents) to be summarized, and define a topic description that will be used to generate topical summaries. Currently the single-document summarizer, and therefore XDOCTOOL, can handle a variety of document formats, including plain ASCII text, HTML format, SGML format, and several other formats used by various news sources.



Clustering Threshold: 0.25

Document Clusters: 10

wisegb001(size=10, threshold=0.25)
CLUSTER1
4 (FT941-4093)
10 (FT942-5119)
6 (FT942-5129)
CLUSTER2
4 (FT941-4093)
5 (FT934-796)
30 (FT941-11985)
CLUSTER3
3 (FT934-398)
5 (FT934-796)
30 (FT941-11985)

Single Documents: 16

19 (FBIS3-18993)
13 (FT942-3893)
22 (FT931-14163)

Cross Document Summary

Number of clusters: 10
Pick Threshold: 0.25

CLUSTERS: 1, 2, 4
Summary FT941-4093
Finance and the Family: Euro Disney deal 'nearer' - At a glance

Beleaguered Euro Disney this week edged closer to a Pounds 1 3bn rescue deal The European leisure group and Walt Disney, its US parent, are understood to have mapped out a restructuring plan with Euro Disney's banks The hope is that an outline rescue can be agreed before the March 31 deadline imposed by Disney for continuing to support Euro Disney The American company is understood to be willing to extend its support beyond the deadline if the banks agree in principle to the rescue package

totalClusters = 26 (10 clusters + 16 singleDocuments)

Example 2. The Multi-Document Tool

The second screen (Example 2) shows the results of the summarization. The large window in the middle of the screen, the text window, provides the user with one of several possible views of the multi-document summary. The possible views include the highest level final MD summary, "cluster" reports which simply are the representative summaries corresponding to each of the clusters (together with information such as common terms, query terms and headers of all documents in the cluster), the single-document summaries for each document in a cluster, and the original document for each document in a cluster. The left window shows a tree structure describing how documents were clustered. Clicking on different 'nodes' in this tree gives the user access to all relevant data, ranging from the final high-level summary to the individual documents. The document numbers corresponding to each document make it easier for the user to keep track, especially in cases when documents are placed in several different clusters.

An important feature of this screen is the slider that represents the threshold mentioned in 'Group Summaries' (3.1.2) used for clustering the documents. If this threshold is high, documents need to have a high degree of similarity to be put in the same cluster. Therefore, the clusters that are formed are smaller in size but more in number (as the selection criteria for membership to a cluster is stricter with a higher threshold). In general, increasing the threshold will result in more clusters, while lowering the threshold will result in fewer clusters. The desired threshold depends on the similarity or dissimilarity of the document collection, the user's preference for high-level topic clusters or sub-topic clusters and the task at hand. The user can change this threshold easily using the slider. Currently, a default threshold is computed based on the 'connectivity' of the collection as implied by the similarity matrix.

4. Evaluation

4.1. Tasks and Evaluation Sets

A formal evaluation of single-document summarizers, SUMMAC (Firmin and Chrzanowski, 1999), was conducted under DARPA's Tipster Text Program (TIPSTER 1998b). This evaluation proved to be a formidable task requiring human judges to manually evaluate the summaries. With the knowledge that the multi-document summarization task is more complex than the single-document one, and given the difficulty faced earlier by the SUMMAC organizers, we wanted to keep the evaluation of our multi-document summarization system as simple as possible. We describe below two different evaluations that we designed for the evaluation.

Initially, we decided to focus entirely on the quality of the multi-document summaries. In other words, we wanted to show that one could quickly and easily massage the multi-document summary produced using our XDOCTOOL into a desired final form. The baseline we wanted to compare this to was producing the desired final multi-document summary by concatenating and massaging the summaries produced as the output of our single-document summarizer.

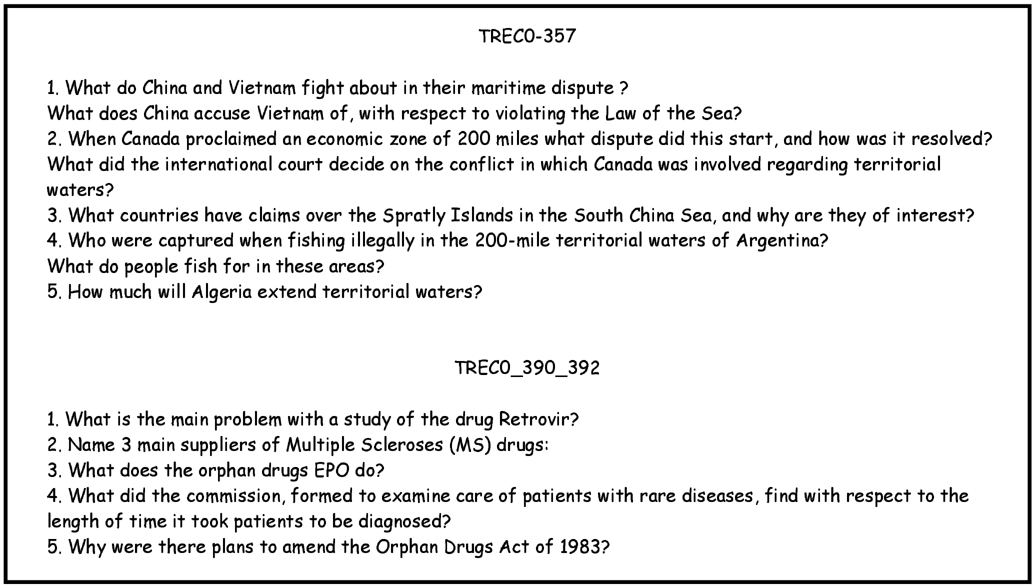
We constructed two different test sets for the evaluation. The first consisted of 60 documents retrieved using Cornell's SMART text retrieval system using TREC topic 357 as the query. Topic 357 deals with articles regarding "territorial water disputes" (TREC-7). The documents returned by SMART are from TREC's document collection. It should be noted that 30 of the 60 documents retrieved by SMART are actually relevant to the query because SMART is not perfect. In this case, topic 357 was chosen primarily because it had at least 30 documents which were identified by the TREC annotators to be relevant. The second set also consists of 60 documents retrieved by SMART. However, unlike the first set, this one consists of 30 relevant articles corresponding to each of the following two TREC topics: topic 390 (orphan drugs), and topic 392 (robotics). Since this second set contains articles from two different TREC topics, we feel that it is easier than the first since it is easier for the user to distinguish between relevant and non-relevant articles for each topic. In comparison, for the first set, the non-relevant documents are quite similar in nature to the relevant ones. Once again, the topics were chosen based on the fact that there were at least 30 documents correcty relevant to each. Please note that it was *not* known how both systems would handle these topics.

For our initial experiment, we had three users use our XDOCTOOL to produce the final multi-document summary while having three others use the single-document summarizer. The final multi-document summary was to contain the main points relevant to the topic of interest. The topic of interest for the second test set was topic 390 (orphan drugs). In addition, we also limited the time each user had to produce the final summary to 15 minutes. There were two main reasons for this limitation:

1. We wanted to mimic a real world scenario where an analyst does not have enough adequate time to read all the documents.

2. Our goal was to analyze the difference in the final summaries produced by the two different summarizers keeping all other factors the same.

Each user produced two final multi-document summaries, one for each of the test sets. The results clearly showed, for both the test sets, that the final summaries created as a result of using the XDOCTOOL were more substantive in terms of the presence of the number of paragraphs corresponding to relevant documents. Since we knew the relevancy judgement for each article in the test sets, computing this was easy. However, we wanted to better quantify the benefit of using the XDOCTOOL. Therefore, we designed a second set of experiments that provided more quantitative data.



TRECO-357

1. What do China and Vietnam fight about in their maritime dispute ?
What does China accuse Vietnam of, with respect to violating the Law of the Sea?
2. When Canada proclaimed an economic zone of 200 miles what dispute did this start, and how was it resolved?
What did the international court decide on the conflict in which Canada was involved regarding territorial waters?
3. What countries have claims over the Spratly Islands in the South China Sea, and why are they of interest?
4. Who were captured when fishing illegally in the 200-mile territorial waters of Argentina?
What do people fish for in these areas?
5. How much will Algeria extend territorial waters?

TRECO_390_392

1. What is the main problem with a study of the drug Retrovir?
2. Name 3 main suppliers of Multiple Scleroses (MS) drugs:
3. What does the orphan drugs EPO do?
4. What did the commission, formed to examine care of patients with rare diseases, find with respect to the length of time it took patients to be diagnosed?
5. Why were there plans to amend the Orphan Drugs Act of 1983?

Example 3. Questions for question-answering task

The second set of experiments was designed in the form of a question answering task. For each of the two test sets described earlier, we constructed a set of questions that we wanted answered. The list of questions for both these tasks is shown in Example 3. These questions were formulated by a person who had never seen the texts before and did not know how the systems would process them. Questions were meant to refer to detailed information in the texts. Similar to the first set of experiments, we had six users, three of whom answered the set of questions for both the test sets using the XDOCTOOL while the other three answered the questions using the single-document summarizer. Our hypothesis here was that using the XDOCTOOL would simplify finding the answers since, for each cluster, the cluster summaries would indicate the presence of the answer in that cluster. Table 1 shows the figures obtained for each user on the two test sets. It should be noted that the users testing the summarization systems had varied amounts of prior experience using the tool. About half the users were novices who had not used the tool before. Hence, the variation in the amount of time among users for each test set.

User	Summarizer	# of questions answered (Test set 1)	Time taken – mins (Test set 1)	# of questions answered (Test set 2)	Time taken – mins (Test set 2)	Average time per test set – mins
1	XDOCTOOL	5/5	18	5/5	15	16.5
2	XDOCTOOL	5/5	27	5/5	26	26.5
3	XDOCTOOL	5/5	11	5/5	12	11.5
4	Single Doc	5/5	24	4/5	32	28
5	Single Doc	5/5	43	4/5	35	39
6	Single Doc	5/5	27	3/5	23	25

Table 1. Results for Question-Answering Task

4.2. Results and User Feedback

The results verify our hypothesis that the XDOCTOOL helps the users find the answers. The numbers in Table 1 clearly show that users using the single-document summarizer took longer to find the answers. In addition, three of them either extracted the incorrect answer or gave up while looking for answers to a few questions.

For the question-answering task, all users of both tools mentioned the need for a keyword search. Some users wanted the possibility of tagging files, so they could keep track of files already investigated.

As expected, the users of the single-document tool had the most complaints; they would like clustering, ranking of documents and organization based on temporal data.

The users of XDOCTOOL liked the cluster reports on which they based their decision to further investigate a cluster. The highlighted terms, indicating terms that documents in a cluster have in common, and terms documents have in common with the topic description, were helpful for quickly scanning the summaries and documents. The top-level summary, however, was not used for the question task. One user pointed out that the clustering should be more user-interactive, for instance, using question descriptions, or assigned weights to terms, to adapt clustering to current needs.

5. Related Work

Uramoto and Takeda (1998) have developed a system that visualizes certain characteristics of a set of documents by organizing them in a directed graph. Although no readable summary is generated, keywords indicated how documents are similar or different. Mani and Bloedorn (1997) also relate pairs of documents to each other showing similarities and differences. In addition, work by McKeown and Radev (McKeown and Radev 1995; Radev and McKeown 1998) relies on an ‘assumed’ system filling and selecting predefined templates used for the final summary. Later work by McKeown et al. (1999) breaks documents into paragraph-based units. These units are compared to each other to identify similar and dissimilar passages. A graph-based one-pass clustering algorithm is applied, using the similarity metric, to identify common topics/themes. Instead of picking a representative sentence from the paragraphs in a cluster, common phrases are identified which are used to generate a new representative sentence. The Carnegie Group’s work on multi-document summarization (Carbonell and Goldstein 1998) relies on the maximal marginal relevance measure to organize the final summary and detect redundancy. Similarly to our approach, clusters are formed and a representative segment is presented to the user.

We are currently not aware of any formal evaluation of multi-document summarization other than work by described in McKeown et al. (1999) and Stein et al. (2000). However, both evaluations are system specific and evaluate certain aspects. The first system focuses on three system components: the similarity metric (evaluated using TDT data), the theme phrase detection approach and the sentence generation capability. The second system analyses the clustering capability and the ability to select representative passages from a cluster. Both evaluations do not compare to a base-line system, nor are they evaluated with respect to certain tasks.

6. Discussion

It is a non-trivial task to carry out a meaningful evaluation for multi-document summarization. First, user subjectivity can be a problem when making judgements about document relevancy and whether something is a main point for a final summary, thus making makes it harder to compare results across users. Our earlier task of producing a final topical summary was affected by these problems. The advantage of the question-answering task is that the final results can be judged objectively. Based on our findings we think that a task-specific user evaluation, particularly the question-answering evaluation, is not only a good way to evaluate a specific multi-document system, but also allows for a system-independent evaluation.

Our experiments showed that the multi-summarization tool helps the user finding specific answers to detailed questions in shorter time than the base-line system.

Acknowledgements

This paper is based upon work supported in part by the Defense Advanced Research Projects Agency under Tipster Phase-3 Contract 97-F157200-000. For help with the experiments we like to thank William Cheetham, Yu-to Chen, Daniel Cleary, Virginia Didomizio, Abha Moitra, and Nam Park.

References

ALLAN, JAMES, JAIME CARBONELL, GEORGE DODDINGTON, JON YAMRON and Y. YANG. 1998. Topic Detection and Tracking Pilot Study: Final Report. In Proceedings of the Broadcast News Understanding and Transcription Workshop, 194 – 218.

CARBONELL, JAIME AND JADE GOLDSTEIN. 1998. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. SIGIR ’98.

EL-HAMDOUCHI, A., and P. WILLETT. 1987. Techniques for the Measurement of Clustering Tendency in Document Retrieval Systems. in Information Science, 13, 361- 65.

EVERITT, BRIAN S. 1993. Cluster Analysis. 3rd Edition. Edward Arnold: London.

FIRMIN, THERESE and MICHAEL J.CHRZANOWSKI. 1999. An Evaluation of Automatic Text Summarization Systems. In Advances in automatic text summarization, Inderjeet Mani and Mark T. Maybury (eds). The MIT Press: Cambridge, Massachusetts.

LORR, M. 1983. Cluster Analysis for Social Scientists: Techniques for Analyzing and Simplifying Complex Blocks of Data. Jossey-Bass: San Francisco.

MANI, I., D. HOUSE, G. KLEIN, L. HIRSCHMAN, L. OBRST, T. FIRMIN, M. CHRZANOWSKI, B. SUNDHEIM. 1998. The TIPSTER SUMMAC Text Summarization Evaluation, Final Report. Mitre Technical Report MTR 98W0000138.

MANI, INDERJEET and ERIC BLOEDORN. 1997. Multi-document Summarization by Graph Search and Matching. AAAI '97, 622-628.

MCKEOWN, KATHLEEN and DRAGOMIR R. RADEV. 1995. Generating Summaries of Multiple News Articles, SIGIR '95, 74-82.

MCKEOWN, KATHLEEN R., JUDITH L. KLAVANS, REGINA BARZILAY and ELEAZAR ESKIN. 1999. Towards Multidocument Summarization by Reformulation: Progress and Prospects. AAAI '99, 453 – 460.

RADEV, DRAGOMIR R., and KATHLEEN R. MCKEOWN. 1998. Generating Natural Language Summaries from Multiple On-Line Sources. Computational Linguistics, Volume **24**, Number 3.

RASMUSSEN, EDIE. 1992. Clustering Algorithms. In Information Retrieval, Data Structures & Algorithms, William B. Frakes, Ricardo Baeza-Yates (editors), Prentice Hall: Englewood Cliffs, New Jersey.

RIJSBERGEN, C. J. VAN. 1979. Information Retrieval. Butterworths: London.

SALTON, GERARD. 1989. Automatic Text Processing. Addison-Wesley Publishing Company: Reading, Massachusetts.

STEIN, GEES C., TOMEK STRZALKOWSKI, G. BOWDEN WISE. 2000, to appear. Interactive, Text-based Summarization of Multiple Documents, Computational Intelligence, 16(4)

STEIN, GEES C., TOMEK STRZALKOWSKI, G. BOWDEN WISE. 1999. Summarizing Multiple Documents using Text Extraction and Interactive Clustering. PACLING '99, August 25-28, University of Waterloo, Ontario.

STEIN, GEES C., TOMEK STRZALKOWSKI, G. BOWDEN WISE AND AMIT BAGGA. 2000. Evaluating Summaries for Multiple Documents in an Interactive Environment. LREC, May 2000.

STRZALKOWSKI, TOMEK, GEES STEIN, JIN WANG AND BOWDEN WISE.1999. A Robust Practical Text Summarizer. In Advances in Automated Text Summarization, Inderjeet Mani and Mark T. Maybury (eds.). The MIT Press.

STRZALKOWSKI, TOMEK, GEES STEIN AND G. BOWDEN WISE. 1998. A Text-Extraction Based Summarizer. TIPSTER Workshop, October 1998.

STRZALKOWSKI, TOMEK, J. WANG AND B. WISE. 1998. Summarization-based Query Expansion in Information Retrieval. COLING-ACL '98, 1258-1264.

TIPSTER 1998a. Tipster text phase III 18-month workshop notes, May, Fairfax, VA.

TIPSTER 1998b. Tipster text phase III 24-month workshop notes, October, Baltimore, MD.

TREC-7. The Seventh Text Retrieval Conference. E.M. VOORHEES and D.K. HARMAN (editors). NIST Special Publication 500-242. Department of Commerce, National Institute of Standards and Technology.

URAMOTO, NAOHIKO and KOICHI TAKEDA. 1998. A Method for Relating Multiple Newspaper Articles by Using Graphs, and Its Application to Webcasting. COLING-ACL '98, 1307-1313.

VOORHEES, E.M. 1986. The Effectiveness and Efficiency of Agglomerative Hierarchic Clustering in Document Retrieval. Ph.D. thesis, Cornell University.