Evaluating automatically generated user-focused multi-document summaries for geo-referenced images

Ahmet Aker
Department of Computer Science
University of Sheffield
Sheffield, S1 4DP, UK
A.Aker@dcs.shef.ac.uk

Robert Gaizauskas
Department of Computer Science
University of Sheffield
Sheffield, S1 4DP, UK
R.Gaizauskas@dcs.shef.ac.uk

Abstract

This paper reports an initial study that aims to assess the viability of a state-of-the-art multi-document summarizer for automatic captioning of geo-referenced images. The automatic captioning procedure requires summarizing multiple web documents that contain information related to images' location. We use SUMMA (Saggion and Gaizauskas, 2005) to generate generic and query-based multi-document summaries and evaluate them using ROUGE evaluation metrics (Lin, 2004) relative to human generated summaries. Results show that, even though query-based summaries perform better than generic ones, they are still not selecting the information that human participants do. In particular, the areas of interest that human summaries display (history, travel information, etc.) are not contained in the query-based summaries. For our future work in automatic image captioning this result suggests that developing the query-based summarizer further and biasing it to account for user-specific requirements will prove worthwhile.

1 Introduction

Retrieving textual information related to a location shown in an image has many potential applications. It could help users gain quick access to the information they seek about a place of interest just by taking its picture. Such textual information could also, for instance, be used by a journalist who is planning to write an article about a building, or by a tourist who seeks further interesting places to visit nearby. In this paper we aim to generate such textual information automatically by utilizing multi-document summarization techniques, where documents to be summarized are web documents that contain information related to the image content. We focus on geo-referenced images, i.e. images tagged with coordinates (latitude and longitude) and compass information, that show things with fixed locations (e.g. buildings, mountains, etc.).

Attempts towards automatic generation of image-related textual information or captions have been previously reported. Deschacht and Moens (2007) and Mori et al. (2000) generate image captions automatically by analyzing image-related text from the immediate context of the image, i.e. existing image captions, surrounding text in HTML documents, text contained in the image, etc. The authors identify named entities and other noun phrases in the image-related text and assign these to the image as captions. Other approaches create image captions by taking into consideration image features as well as image-related text (Westerveld, 2000; Barnard et al., 2003; Pan et al., 2004). These approaches can address all kinds of images, but focus mostly on images of people. They analyze only the immediate textual context of the image on the web and are concerned with describing *what* is in the image only. Consequently, background information about the objects in the image is not provided. Our aim, however, is to have captions that inform users' specific interests about a location, which clearly includes more than just image content description. Multi-document summarization techniques offer the possibility to include image-related information from multiple

documents, however, the challenge lies in being able to summarize unrestricted web documents.

Various multi-document summarization tools have been developed: SUMMA (Saggion and Gaizauskas, 2005), MEAD (Radev et al., 2004), CLASSY (Conroy et al., 2005), CATS (Farzinder et al., 2005) and the system of Boros et al. (2001), to name just a few. These systems generate either generic or query-based summaries or both. Generic summaries address a broad readership whereas query-based summaries are preferred by specific groups of people aiming for quick knowledge gain about specific topics (Mani, 2001). SUMMA and MEAD generate both generic and query-based multi-document summaries. Boros et al. (2001) create only generic summaries, while CLASSY and CATS create only query-based summaries from multiple documents. The performance of these tools has been reported for DUC tasks[1]. As Sekine and Nobata (2003) note, although DUC tasks provide a common evaluation standard, they are restricted in topic and are somewhat idealized. For our purposes the summarizer needs to create summaries from unrestricted web input, for which there are no previous performance reports.

For this reason we evaluate the performance of both a generic and a query-based summarizer and use SUMMA which provides both summarization modes. We hypothesize that a query-based summarizer will better address the problem of creating summaries tailored to users' needs. This is because the query itself may contain important hints as to what the user is interested in. A generic summarizer generates summaries based on the topics it observes from the documents and cannot take user specific input into consideration. Using SUMMA, we generate both generic and query-based multi-document summaries of image-related documents obtained from the web. In an online data collection procedure we presented a set of images with related web documents to human subjects and asked them to select from these documents the information that best describes the image. Based on this user information we created model summaries against which we evaluated the automatically generated ones.

Section 2 in this paper describes how image-related documents were collected from the web. In section 3 SUMMA is described in detail. In section 4 we explain how the human image descriptions were collected. Section 5 discusses the results, and section 6 concludes the paper and outlines directions for future work and improvements.

2 Web Document Collection

For web document collection we used geo-referenced images of locations in London such as *Westminster Abbey, London Eye, etc.* The images were taken with a digital SLR camera with a Geotagger plugged-in to its flash slot. The Geotagger helped us to identify the location by means of coordinates of the position where the photographer stands, as well as the direction the camera is pointing (compass information). Based on the coordinates and compass information for each image, we carried out the following steps to collect related documents from the web:

- identify a set of toponyms (terms that denote locations or associate names with locations, e.g. *Westminster Abbey*) that can be passed to a search engine as query terms for document search;
- use a search engine to retrieve HTML documents to be summarized;
- extract the pure text out of the HTML documents.

2.1 Toponym Collection

In order to create the web queries a set of toponyms were collected semi-automatically. We implemented an application (cf. Figure 1) that suggests a list of toponyms close to the photographer's location. The application uses Microsoft's MapPoint[2] service which allows users to query location-related information. For example, a user can query for tourist attractions (interesting buildings, museums, art galleries etc.) close to a location that is identified by its address or its coordinates.

Based on the coordinates (latitude and longitude), important toponyms for a particular image can be queried from the MapPoint database. In order to facilitate this, MapPoint returns a metric that measures the importance of each toponym. A value close to zero means that the returned toponym is closer to the specified coordinates than a toponym with a higher value. For instance for

[1] http://www-nlpir.nist.gov/projects/duc/index.html

[2] http://www.microsoft.com/mappoint/

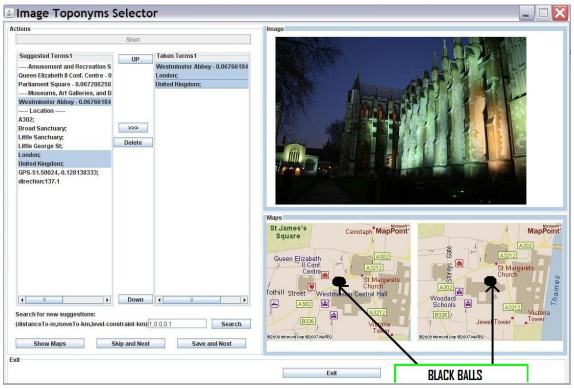

Figure 1: Image Toponym Collector: Westminster Abbey, Lat: 51.50024 Lon: -0.128138333: Direction: 137.1

the image of *Westminster Abbey* shown in the *Image* box of Figure 1 the following toponyms are collected:

```
Queen Elizabeth II Conf. Centre: 0.059
Parliament Square: 0.067
Westminster Abbey: 0.067
```

The photographer's location is shown with a black dot on the first map in the *Maps* box of Figure 1. The application suggests the toponyms shown in the *Suggested Terms* list.

Knowing the direction the photographer was facing helps us to select the correct toponyms from the list of suggested toponyms. The current MapPoint implementation does not allow an arrow to be drawn on the map which would be the best indication of the direction the photographer is facing. To overcome this problem we create a second map (cf. *Maps* box of Figure 1) that shows another dot moved 50 meters in the compass direction. By following the dot from the first map to the second map we can determine the direction the photographer is facing. When the direction is known, it is certain that the image shows *Westminster Abbey* and not the *Queen Elizabeth II Conf. Centre* or *Parliament Square*. The *Queen Elizabeth II Conf. Centre* is behind the photographer and *Parliament Square* is on the left hand side.

Consequently in this example the toponym *Westminster Abbey* is selected manually for the web search. In order to avoid ambiguities, the city name and the country name (also generated by MapPoint) are added manually to the selected toponyms. Hence, for *Westminster Abbey*, *London* and *United Kingdom* are added to the toponym list. Finally the terms in the toponym list are simply separated by a boolean *AND* operator to form the web query. Then, the query is passed to the search engine as described in the next section.

2.2 Document Query and Text Extraction

The web queries were passed to the Google Search engine and the 20 best search results were retrieved, from which only 11 were taken for the summarization process. We ensure that these 20 search results are healthy hyperlinks, i.e. that the content of the hyperlink is accessible. In addition to this, multiple hyperlinks belonging to the same domain are ignored as it is assumed that the content obtained from the same domain would be similar. Each remaining search result is crawled to obtain its content.

The web-crawler downloads only the content of the document residing under the hyperlink, which was previously found as a search result, and does not follow any other hyperlinks within the document. The content obtained by the web-crawler encapsulates an HTML structured document. We further process this using an HTML parser[3] to select the *pure* text, i.e. text consisting of sentences.

The HTML parser removes advertisements, menu items, tables, java scripts etc. from the HTML documents and keeps sentences which contain at least 4 words. This number was chosen after several experiments. The resulting data is passed on to the multi-document summarizer which is described in the next section.

3 SUMMA

SUMMA[4] is a set of language and processing resources to create and evaluate summarization systems (single document, multi-document, multilingual). The components can be used within GATE[5] to produce ready summarization applications. SUMMA has been used in this work to create an extractive multi-document summarizer: both generic and query-based.

In the case of generic summarization SUMMA uses a single cluster approach to summarize *n* related documents which are given as input. Using GATE, SUMMA first applies sentence detection and sentence tokenisation to the given documents. Then each sentence in the documents is represented as a vector in a vector space model (Salton, 1988), where each vector position contains a term

[3] http://htmlparser.sourceforge.net/
[4] http://www.dcs.shef.ac.uk/ saggion/summa/default.htm
[5] http://gate.ac.uk

(word) and a value which is a product of the *term frequency* in the document and the *inverse document frequency (IDF)*, a measurement of the term's distribution over the set of documents (Salton and Buckley, 1988). Furthermore, SUMMA enhances the sentence vector representation with further features such as the sentence position in its document and the sentence similarity to the lead-part in its document. In addition to computing the vector representation for all sentences in the document collection the centroid of this sentence representation is also computed.

In the sentence selection process, each sentence in the collection is ranked individually, and the top sentences are chosen to build up the final summary. The ranking of a sentence depends on its distance to the centroid, its absolute position in its document and its similarity to the lead-part of its document. For calculating vector similarities, the cosine similarity measure is used (Salton and Lesk, 1968).

In the case of the query-based approach, SUMMA adds an additional feature to the sentence vector representation as computed for generic summarization. For each sentence, cosine similarity to the given query is computed and added to the sentence vector representation. Finally, the sentences are scored by summing all features in the vector space model according to the following formula:

$$Sentence_{score} = \sum_{i=1}^{n} feature_i * weight_i$$

After the scoring process, SUMMA starts selecting sentences for summary generation. In both generic and query-based summarization, the summary is constructed by first selecting the sentence that has the highest score, followed by the next sentence with the second highest score until the compression rate is reached. However, before a sentence is selected a similarity metric for redundancy detection is applied to each sentence which decides whether a sentence is distinct enough from already selected sentences to be included in the summary or not. SUMMA uses the following formula to compute the similarity between two sentences:

$$NGramSim(S_1, S_2, n) = \sum_{j=1}^{n} w_j * \frac{grams(S_1, j) \bigcap grams(S_2, j)}{grams(S_1, j) \bigcup grams(S_2, j)}$$

where n specifies maximum size of the n-grams to be considered, grams(S_X, j) is the set of j-grams in sentence X and w_j is the weight associated with j-gram similarity. Two sentences are similar if *NGramSim(S_1, S_2, n)* $> \alpha$. In this work n is set to 4 and α to 0.1. For j-gram similarity weights $w_1 = 0.1$, $w_2 = 0.2$, $w_3 = 0.3$ and $w_4 = 0.4$ are selected. These values are coded in SUMMA as defaults.

Using SUMMA, generic and query-based summaries are generated for the image-related documents obtained from the web. Each summary contains a maximum of 200 words. The queries used in the query-based mode are toponyms collected as described in section 2.1.

4 Creating Model Summaries

For evaluating automatically generated summaries as image captions, information that people associate with images is collected. For this purpose, an online data collection procedure was set up. Participants were provided with a set of 24 images. Each image had a detailed map showing the location where it was taken, along with URLs to 11 related documents which were used for the automated summarization. Figure 2 shows an example of an image and Table 2 contains the corresponding related information.

Each participant was asked to familiarize him- or herself with the location of the image by analyzing the map and going through all 11 URLs. Then each participant decided on up to 5 different pieces of information he/she would like to know if he/she sees the image or information about something he/she relates with the image. The information we collected in this way is similar to 'information nuggets' (Voorhees, 2003). Information nuggets are facts which help us assess automatic summaries by checking whether the summary contains the fact or not. In addition to this, each participant was asked to collect the information only from the given documents, ignoring any other links in these documents.

Eleven students participated in this survey, simulating the scenario in which tourists look for information about an image of a popular sight. The number of images annotated by each participant is shown in Table 1.

The participants selected the information from original HTML documents on the web and not from the documents which were preprocessed for the multi-document summarization task. We found

Table 1: Number of images annotated by each particant

User1	User2	User3	User4	User5	User6	User7	User8	User9	User10	User11
24	7	24	24	18	24	8	4	16	12	24

Figure 2: Example image

Table 2: Information related to Figure 2

1. Westminster Abbey is the place of the coronation, marriage and burial of British monarchs, except Edward V and Edward VIII since 1066
2. the parish church of the Royal Family
3. the centrepiece to the City of Westminster
4. first church on the site is believed to have been constructed around the year 700
5. The history and the monuments, crypts and memorials are not to be missed.

out that in some cases the participants selected information that did not occur in the preprocessed documents. To ensure that the information selected by the participants also occurs in the preprocessed documents, we retained only the information selected by the participants that could also be found in these documents, i.e. that was available to the summarizer. Out of 807 nuggets selected by participants 21 (2.6%) were not found in the documents available to the summarizer and were removed.

Furthermore, as the example above shows (cf. Table 2), not all the items of information selected by the participants were in form of full sentences. They vary from phrases to whole sentences. The participants were free to select any text unit from the documents that they related to the image content. However, SUMMA works extractively and its summaries contain only sentences selected from the given input documents. The user selected information was normalized to sentences in order to have comparable summaries for evaluation. This was achieved by selecting the sentence(s) from the documents in which the participant-selected information was found and replacing the participant-selected phrases or clauses with the full sentence(s). In this way model summaries were obtained.

5 Results

The model summaries were compared against 24 summaries generated automatically using SUMMA by calculating ROUGE-1 to ROUGE-4, ROUGE-L and ROUGE-W-1.2 recall metrics (Lin, 2004). For all these metrics ROUGE compares each automatically generated summary s pairwise to every model summary m_i from the set of M model summaries and takes the maximum $ROUGE_{Score}$ value among all pairwise comparisons as the best $ROUGE_{Score}$ score:

$$ROUGE_{Score} = argmax_i ROUGE_{Score}(m_i, s)$$

ROUGE repeats this comparison M times. In each iteration it applies the Jackknife method and takes one model summary from the M model summaries away and compares the automatically generated summary s against the $M-1$ model summaries. In each iteration one best $ROUGE_{Score}$ is calculated. The final $ROUGE_{Score}$ is then the average of all best scores calculated in M iterations.

In this way each generic and query-based summary was compared with the corresponding model summaries. The results are given in the first two columns of Table 3. We also collected the common information all participants selected for a particular image and compared this to the corresponding query-based summary. The common information is the intersection set of the sets of information each of the participants selected for a particular image. The results for this comparison are shown in column *QueryToCPOfModel* of Table 3.

The model summaries were also compared against each other in order to assess the agreement between the participants. To achieve this, the image information selected by each participant was compared against the rest. The corresponding results are shown in column *UserToUser* of Table 4. We applied the same pairwise comparison we used for our model summaries to the model summaries of task 5 in DUC 2004 in order to mea-

Table 3: Comparison: Automatically generated summaries against model summaries. The column GenericToModel for example shows ROUGE results for generic summaries relative to model summaries. CP stands for common part, i.e. common information selected by all participants.

Recall	GenericToModel	QueryToModel	QueryToCPOfModel	QueryToModelInDUC
R-1	0.38293	0.39655	0.22084	0.3341
R-2	0.14760	0.17266	0.09894	0.0723
R-3	0.09286	0.11196	0.06222	0.0279
R-4	0.07450	0.09219	0.04971	0.0131
R-L	0.34437	0.35837	0.20913	0.3320
R-W-1.2	0.11821	0.12606	0.06350	0.1130

Table 4: Comparison: Model summaries against each other

Recall	UserToUser	UserToUserInDUC
R-1	0.42765	0.45407
R-2	0.30091	0.13820
R-3	0.26338	0.05870
R-4	0.24964	0.02950
R-L	0.40403	0.41594
R-W-1.2	0.15846	0.13973

sure the agreements between the participants on this standard task. This gives us a benchmark relative to which we can assess how well users agree on what information should be related to images. The results for this comparison are shown in column *UserToUserInDUC* of Table 4.

All ROUGE metrics except R-1 and R-L indicate higher agreement in human image-related summaries than in DUC document summaries. The ROUGE metrics most indicative of agreement between human summaries are those that best capture words occurring in longer sequences of words immediately following each other (R-2, R-3, R-4 and R-W). If long word sequences are identical in two summaries it is more likely that they belong to the same sentence than if only single words are common, as captured by R-1, or sequences of words that do not immediately follow each other, as captured by R-L. In R-L gaps in word sequences are ignored so that for instance *A B C D G* and *A E B F C K D* have the common sequence *A B C D* according to R-L. R-W considers the gaps in words sequences so that this sequence would not be recognized as common. Therefore the agreement on our image-related human summaries is substantially higher than agreement on DUC document human summaries.

The results in Table 3 support our hypothesis that query-based summaries will perform better than generic ones on image-related summaries. All ROUGE results of the query-based summaries are greater than the generic summary scores. This reinforces our decision to focus on query-based summaries in order to create image-related summaries which also satisfy the users' needs. However, even though the query-based summaries are more appropriate for our purposes, they are not completely satisfactory. The query-based summaries cover only 39% of the unigrams (ROUGE 1) in the model summaries and only 17% of the bigrams (ROUGE 2), while the model summaries have 42% agreement in unigrams and 30% agreement in bigrams (cf. column *UserToUser* in Table 4). The agreement between the query-based and model summaries gets lower for ROUGE-3 and ROUGE-4 indicating that the query-based summaries contain very little information in common with the participants' results. This indication is supported by the ROUGE-L (35%) and the low ROUGE-W (12%) agreement which are substantially lower compared to the *UserToUser* ROUGE-L (40%) and ROUGE-W (15%) and the low ROUGE scores in column *QueryToCPOfModel*. For comparison with automated summaries in a different domain, we include ROUGE scores of query based SUMMA used in DUC 2004 (Saggion and Gaizauskas, 2005) as shown in the last column of Table 3. All scores are lower than our *QueryToModel* results which might be due to low agreement between human generated summaries for the DUC task (cf. *UserToUserInDUC* column in Table 4) or maybe because image captioning is an easier task. The possibility that our summarization task is easier than DUC due to the summarizer having fewer documents to summarize or due to the documents being shorter than those in the DUC task can be excluded. In the DUC task the multi-document clusters contain 10 documents on average while our summarizer works with 11 documents. The mean length in documents in DUC

Table 5: Query-based summary for Westminster Abbey and information selected by participants

Query-based summary	Information selected by participants
The City of London has St Pauls, but Westminster Abbey is the centrepiece to the City of Westminster. Westminster Abbey should be at the top of any London traveler's list. Westminster Abbey, however, lacks the clear lines of a Rayonnant church,... I loved Westminster Abbey on my trip to London. **Westminster Abbey was rebuilt after 1245 by Henry III's order, and in 1258 the remodeling of the east end of St. Paul's Cathedral began.** He was interred in Westminster Abbey. From 1674 to 1678 he tuned the organ at Westminster Abbey and was employed there in 1675-76 to copy organ parts of anthems. The architectural carving found at Westminster Abbey (mainly of the 1250s) has much of the daintiness of contemporary French work, although the drapery is still more like that of the early Chartres or Wells sculpture than that of the Joseph Master. Nevertheless, Westminster Abbey is something to see if you have not seen it before. I happened upon the Westminster Abbey on an outing to Parliament and Big Ben.	**1.***(3)* Westminster Abbey is the place of the coronation, marriage and burial of British monarchs, except Edward V and Edward VIII since 1066. **2.***(1)* **What is unknown, however is just how old it is. The first church on the site is believed to have been constructed around the year 700**. **3.***(2)* Standing as it does between Westminster Abbey and the Houses of Parliament, and commonly called "the parish church of the House of Commons", St Margaret's has witnessed many important events in the life of this country. **4.***(1)* In addition, the Abbey is the parish church of the Royal Family, when in residence at Buckingham Palace. **5.***(1)* The history and the monuments, crypts and memorials are not to be missed. **6.***(1)* For almost one thousand years, Westminister Abbey has been the setting for much of London's ceremonies such as Royal Weddings, Coronations, and Funeral Services. **7.***(1)* It is also where many visitors pay pilgrimage to The Tomb of the Unknown Soldier. **8.***(1)* The City of London has St Pauls, but Westminster Abbey is the centrepiece to the City of Westminster.

is 23 sentences while our documents have 44 sentences on average.

Table 5 shows an example query-based summary for the image of *Westminster Abbey* and the information participants selected for this particular image. Jointly the participants have selected 8 different pieces of information as indicated by the bold numbers in the table. The numbers in parentheses show the number of times that a particular information unit was selected. By comparing the two sides it can be seen that the query-based summary does not cover most of the information from the list with the exception of item 2. The item 2 is semantically related to the sentence in bold on the summary side as it addresses the year the abbey was built, but the information contained in the two descriptions is different.

Our results have confirmed our hypothesis that query-based summaries will better address the aim of this research, which is to get summaries tailored to users' needs. A generic summary does not take the user query into consideration and generates summaries based on the topics it observes. For a set of documents containing mainly historical and little location-related information, a generic summary will probably contain a higher number of history-related than location-related sentences. This might satisfy a group of people seeking historical information, however, it might not be interesting for a group who want to look for location-related information. Therefore using a query-based multi-document summarizer is more appropriate for image-related summaries than a generic one. However, the results of the query-based summaries show that even so they only cover a small part of the information the users select. One reason for this is that the query-based summarizer takes relevant sentences according to the query given to it and does not take into more general consideration the information likely to be relevant to the user. However, we can assume that users will have shared interests in some of the information they would like to get about a particular type of object in an image (e.g. a bridge, church etc.). This assumption is supported by the high agreement between participants' performances in our online survey (cf. column *UserToUser* of Table 4).

Therefore, one way to improve the performance of the query-based summarizer is to give the summarizer the information that users typically associate with a particular object type as input and bias the multi-document summarizer towards this information. To do this we plan to build models of user preferences for different object types from the large number of existing image captions from web resources, which we believe will improve the quality of automatically generated captions.

6 Conclusion

In this work we showed that query-based summarizers perform slightly better than generic summarizers on an image captioning task. However, their output is not completely satisfactory when compared to what human participants indicated as important in our data collection study. Our future work will concentrate on extending the query-

based summarizer to improve its performance in generating captions that match user expectations regarding specific image types. This will include collecting a large number of existing captions from web sources and applying machine learning techniques for building models of the kinds of information that people use for captioning. Further work also needs to be carried out on improving the readability of the extractive caption summaries.

7 Acknowledgement

This work is supported by the EU-funded TRIPOD project [6]. We would like to thank Horacio Saggion for his support with SUMMA. We are also grateful to Emina Kurtic, Mark Sanderson, Mesude Bicak and Dilan Paranavithana for comments on the previous versions of this paper.

References

Barnard, Kobus and Duygulu, Pinar and Forsyth, David and de Freitas, Nando and Blei M, David and Jordan I, Michael. 2003. *Matching words and pictures*. *The Journal of Machine Learning Research, MIT Press Cambridge, MA, USA*, 3: 1107–1135.

Boros, Endre and Kantor B, Paul and Neu j, David. 2001. *A Clustering Based Approach to Creating Multi-Document Summaries*. *Proc. of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Conroy M, John and Schlesinger D, Judith and Stewart G, Jade 2005. *CLASSY query-based multi-document summarization*. *Proc. of the 2005 Document Understanding Workshop, Boston*.

Deschacht, Koen and Moens F, Marie. 2007. *Text Analysis for Automatic Image Annotation*. *Proc. of the 45th Annual Meeting of the Association for Computational Linguistics, Prague*.

Farzindar, Atefeh and Rozon, Frederik and Lapalme, Guy. 2005. *CATS a topic-oriented multi-document summarization system at DUC 2005*. *Proc. of the 2005 Document Understanding Workshop (DUC2005)*.

Lin, Chin-Yew 2004. *ROUGE: A Package for Automatic Evaluation of Summaries*. *Proc. of the Workshop on Text Summarization Branches Out (WAS 2004)*.

Mani, Inderjeet. 2001. *Automatic Summarization*. *John Benjamins Publishing Company*.

Mori, Yasuhide and Takahashi, Hironobu and Oka, Ryuichi. 2000. *Automatic word assignment to images based on image division and vector quantization*. *Proc. of RIAO 2000: Content-Based Multimedia Information Access*.

[6] http://tripod.shef.ac.uk/

Pan, Jia-Yu. and Yang, Hyung-Jeong and Duygulu, Pinar and Faloutsos, Christos. 2004. *Automatic image captioning*. *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*.

Radev R, Dragomir. and Jing, Hongyan and Styś, Malgorzata and Tam, Daniel. 2004. *Centroid-based summarization of multiple documents*. *Information Processing and Management*,40(6): 919–938.

Saggion, Horacio and Gaizauskas, Robert 2004. *Multi-document summarization by cluster/profile relevance and redundancy removal*. *Document Understanding Conference (DUC04)*.

Salton, Gerhard 1988. *Automatic text processing*. *Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA*.

Salton, Gerhard and Buckley, Chris 1988. *Term-weighting approaches in automatic text retrieval*. *Pergamon Press, Inc. Tarrytown, NY, USA*.

Salton, Gerhard and Lesk E., Michael 1968. *Computer Evaluation of Indexing and Text Processing*. *Journal of the ACM*,15(1):8–36.

Sekine, Satoshi and Nobata, Chikashi. 2003. *A Survey for Multi-Document Summarization*. *Association for Computational Linguistics Morristown, NJ, USA, Proc. of the HLT-NAACL 03 on Text summarization workshop-Volume 5*.

Voorhees M, Ellen. 2003. *Overview of the TREC 2003 Question Answering Track*. *Proc. of the Twelfth Text REtrieval Conference (TREC 2003)*.

Westerveld, Thijs. 2000. *Image retrieval: Content versus context*. *Content-Based Multimedia Information Access, RIAO 2000 Conference*.