

# Using Domain Knowledge for Text Summarization in Medical Domain

Kamal Sarkar

Computer Science & Engineering Department,  
Jadavpur University,  
Kolkata – 700 032, INDIA,  
[jukamal2001@yahoo.com](mailto:jukamal2001@yahoo.com)

**Abstract**—Medical Literatures on the web are the important sources to help clinicians in patient-care. Initially, the clinicians go through the author-written abstracts or summaries available with the medical articles to decide whether articles are suitable for in-depth study. Since all medical articles do not come with author written abstracts or summaries, automatic summarization of medical articles will help clinicians or medical students to find the relevant information on the web rapidly.

In this paper we discuss a summarization method, which combines several domain specific features with some other known features such as term frequency, title and position to improve the summarization performance in the medical domain. Our experiments show that the incorporation of domain specific features improves the summarization performance.

**Index Terms**— Text summarization, Domain-specific features, Novel medical term detection

## I. INTRODUCTION

These days, people are overwhelmed by the huge amount of information on the Web. The number of pages available on the Internet almost doubles every year. This is also the case for medical information [1], which is now available from a variety of sources. Medical Literature such as medical news, research articles, clinical trial reports on the web are the important source to help clinicians in patient treatment. Initially, the clinicians go through the author-written abstracts or summaries available with the medical articles to decide whether articles are relevant to them for in-depth study. Since all medical articles do not come with author written abstracts or summaries, automatic summarization of medical articles will help clinicians or medical students to find the relevant information on the web rapidly. Moreover, monitoring infectious disease outbreaks or other biological threats demand rapid information gatherings and summarization.

Text summarization is the process to produce a condensed representation of the content of its input for human consumption [2]. Input to a summarization process can be one or more text documents. When only one document is the input, it is called single document text summarization and when the input is a cluster of related text documents, it is multi-document summarization. We can also categorize the text summarization based on the type of users the summary is intended for: User focused (query focused) summaries are tailored to the requirements of a particular user or group

users and generic summaries are aimed at a broad readership community [2].

Depending on the nature of text representation in the summary, summary can be categorized as an abstract and an extract. An extract is a summary consisting of a number of salient text units selected from the input. An abstract is a summary, which represents the subject matter of the article with the text units, which are generated by reformulating the salient units selected from the input. An abstract may contain some text units, which are not present in to the input text.

Based on the information content of the summary, it can be categorized as informative and indicative summary. The indicative summary presents an indication about an article's purpose and approach to the user for selecting the article for in-depth reading; informative summary covers all salient information in the document at some level of detail, i.e., it will contain information about all the different aspects such as article's purpose, scope, approach, results and conclusions etc. For example, an abstract of a medical research article is more informative than its headline.

According to the above-mentioned types and sub-types of automatic text summarization, the summarization technique presented in this paper can be called sentence extraction-based single document informative summarization in medical domain.

Some previous works on extractive summarization used few or all of the features such as term frequency, positional information and cue phrases to compute sentence scores [3] [4] [5]. Some machine learning approaches to extractive summarization had already been investigated. In [6] sentence extraction is viewed as a Bayesian classification task.

Compared to creating an extract, generation of abstract is relatively harder since the latter requires: (1) semantic representation of text units (sentences or paragraphs) in the text, (2) reformulation of two or more text units and (3) rendering the new representation in natural language. Abstractive approaches have used template based information extraction, information fusion and compression. In information extraction based approach, predefined template slots are filled with the desired pieces of information extracted by the summarization engine [7]. Jing and McKeown [8] pointed out that human summaries are often constructed from the source document by a process of cutting and pasting document fragments that are then combined and regenerated as

summary sentences. Jing [9] reports on automated techniques to build a corpus representing the cut-and-paste process used by humans so that such a corpus can then be used to train an automated summarizer. True abstraction requires more sophisticated process that requires large-scale resources. Since large-scale resources of this kind are difficult to develop, abstractive summarization has not progressed beyond the proof-of-concept stage.

Most of the researchers extend to the medical domain the summarization techniques already used in other domains. One of the projects in medical domain is MiTAP [10]. MiTAP (MITRE Text and Audio Processing) monitors infectious disease outbreaks or other biological threats by monitoring multiple information sources. The work presented in [11] exploits extractive techniques, which ranks the extracted sentences according to the so-called cluster signature of the document. The abstracts and full texts from the Journal of the American Medical Association were used for their experiments. TRESTLE (Text Retrieval Extraction and Summarization Technologies for Large Enterprises) is a system, which produces single sentence summaries of pharmaceutical newsletters [12]. TRESTLE generates summaries by filling the templates by the Information Extraction process. The system HelpfulMed [13] helps professional and advanced users to access medical information on the Internet and in medical related databases. An ontology based summarization approach has been proposed in [14]. A query based medical information summarization system that exploits ontology knowledge has been proposed in [15]

The work presented in [15] uses ontology to expand query words and assigns scores to sentences based on number of original keywords (query words) and expanded keywords.

Compared to the above-mentioned approaches, we used domain knowledge more efficiently in measuring the sentence importance in generic summarization framework. In generic text summarization only one summary is produced for a text document. In query-focused summarization, the nature of summary varies depending on the user query.

In section II we discuss how to build up domain knowledge. In section III, overall architecture of the system has been discussed. We present the evaluation and the experimental results in section IV.

## II. DOMAIN KNOWLEDGE PREPARATION

In the domain specific text summarization, one of important task is to discover terms, phrases specific to the domain. In this work, we identified a list of cue terms and phrases specific to the medical domain. The idea is that the phrases like "We report", "We present", "World Health Organization", "This study is", "Prevention of" etc. affect the probable summary worthiness of the sentences containing those phrases while summarizing medical scientific or health news articles. We identified a list of 81 medical cue phrases and terms from a corpus of medical news articles and original research articles. The cue phrases and terms are categorized based on its usefulness in the summarization task. Some cue phrases

such as "We report", "This study is" are more important than the phrase "a study was" because the first two phrases indicates about the recent findings presented by the author in a original research article to be summarized whereas the latter indicates the comparison of author's present concepts to some previous works. So, we categorize the cue phrases and terms into the different types and accordingly we assign empirical scores ranging from 1 to 10 to those phrases and terms. Another important issue in this context is the position of a cue phrase in the sentence. If the phrase "The World Health Organization" appears at the beginning of a sentence, it indicates that the sentence carries some information about the guidelines given by the World Health Organization (WHO) in the context of medical issues related to the disease or medication covered in the article. So, the sentence containing a cue phrase at the beginning gets higher score and it gets relatively lower score if it contains the cue phrase in other position. So, during the preparation of knowledge base, we assign the weights ranging from 1 to 8 to cue phrases based on its effectiveness in determining the summary worthiness. The cue phrase and the corresponding weights are stored in the knowledge base. The additional weight in the range 0 to 2 is added to the cue phrase due to its position in the sentence while the summarization process starts. In effect, a cue phrase gets a weight ranges between 1 and 10.

But, it is difficult to identify all domain specific cue phrases and terms by hands. Moreover, new medical terms such as names of genes, medicines and diseases are discovered from time to time. If an article is about the discovery of a gene responsible for a particular disease, it is necessary to detect this information to make the summary more informative. The idea is to identify the novel medical terms in the article. The novel medical terms are the terms, which are not included in any of existing lexicons of technical terms of the domain. We used an algorithm for novel medical term detection. The algorithm uses some surface level features and out of vocabulary feature to identify the novel medical terms. The out-of-vocabulary (OOV) feature specifies the absence of a term in the known vocabulary. To decide whether a term is out of vocabulary, we used two vocabularies. One vocabulary is built up by using MeSH (Medical Subject Headings), which is NLM's (U.S. National Library of Medicine) controlled vocabulary thesaurus. All the MeSH terms are treated as key phrases. The keyword list was automatically generated from the key phrase list. To obtain the keywords, all key phrases (MeSH terms) were split into individual words and added as keywords to the keyword list. We prepare one table for medical keywords.

But, if a word is not found in the medical vocabulary, it should not be treated as a novel word, because the medical article, which is written using a natural language, contains words such as verbs, adverbs and some nouns, which may not be available in the medical vocabulary. So, we used another vocabulary of words, which has been constructed from a corpus of natural language texts (not related to the medical domain). We used the collection of the news documents downloaded from the site under the

heading Yahoo news coverage and built up a vocabulary of English words. If a term is not available in these two vocabularies, we treat the term as novel medical term.

### III. SYSTEM ARCHITECTURE

Our system consists of three primary components: document preprocessing, sentence extraction and summary generation. Document preprocessing component deals with formatting the input document, segmentation and stop removal. Sentence ranking component assigns scores to the sentences based on the domain knowledge, word level and sentence level features. The summary generation component selects top  $n$  sentences based on scores. Finally, the sentences included in to the summary are reordered to increase the readability. The fig. 1 shows the overall system architecture along with its components.

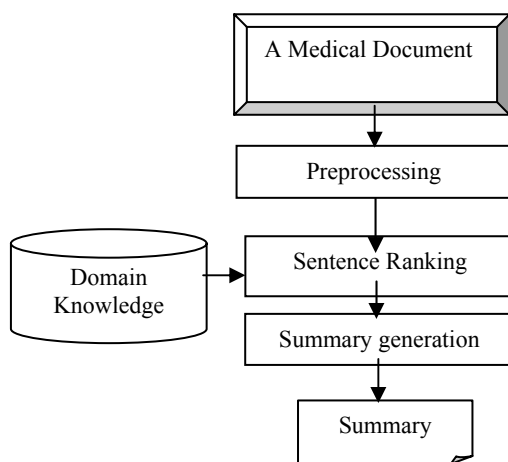


Figure 1. System architecture.

#### A. Document Preprocessing

At this step, input documents are segmented in to sentences. The sentence boundary is identified by a period (.). The preprocessing task primarily includes removal of stop words (prepositions, articles and other low content words), punctuation marks (except dots) and handling of abbreviations. The dots in abbreviations and numeric words (ex. 12.5 millions) may mistakenly be recognized as a sentence boundary. We have used a number of syntactic rules to differentiate between dots in abbreviations, numeric words and dots at the end of the sentence.

We replace the dots in abbreviations by the special character (^). We also remove commas (,) from the numeric words (ex. 12,000 people).

After formatting the input documents, these are now easily broken in to a cluster of sentences. The punctuation marks, parentheses etc. are removed at this step.

In the final summary, we replace the special characters (^) by dots to transform the words in to original form.

#### B. Proposed Sentence Ranking Algorithm

After completion of preprocessing task, we rank the sentences based on their scores. The score of a sentence is

computed based on the many important factors: (1) Frequency of the terms (tf), (2) Similarity of the sentence to the document's title, (3) position of the sentence in the text (4) the presence of domain specific cue phrases in the sentence (5) presence of novel terms in the sentence (6) sentence length. The sentence score is computed as follows:

$$S(s) = S_{tf}(s) + S_{title}(s) + S_{pos}(s) + S_{cue}(s) + S_{novel}(s) + S_{pen}(s), \quad (1)$$

where

$$S_{tf}(s) = \sum_{w \in s} (1 + \log(\text{tf}(w))), \quad (2)$$

$$S_{title}(s) = |s \cap h|, \quad (3)$$

$$S_{pos}(s) = \alpha + 1/\text{pos}(s), \text{ if } \text{pos}(s) < n \\ = 1/\text{pos}(s), \text{ otherwise,} \quad (4)$$

$$S_{cue}(s) = \sum_{c \in s} \text{weight}(c), \quad (5)$$

$$S_{novel}(s) = \sum_{d \in s} \text{weight}(d), \quad (6)$$

$$S_{penalty}(s) = L_{\min}(s) - L(s). \quad (7)$$

$S_{tf}(s)$  = score assigned to a sentence based on the number of higher frequent words it contain.  $\text{tf}(w)$  = number of times a term (or word)  $w$  occurs in the text. Noisy words from the sentences are removed by setting the frequency to some predefined threshold.

$S_{title}(s)$  is the score of the sentence  $S$  due to its similarity to the headline of the document to be summarized.  $|s \cap h|$  represents number of terms common between the sentence  $s$  and the headline  $h$ . This is so-called title feature of Edmundson [4].

$S_{pos}(s)$  is the score assigned to a sentence due to its position in the input text.  $\text{pos}(s)$  indicates the position of the sentence  $s$ . If the position is incremented the importance of the sentences decreases to some degree. The idea behind the positional feature is that author of an article may summarize the main concepts within the first few paragraphs before further elaboration.

$S_{cue}(s)$  =score assigned to a sentence based on the number of cue phrases it contains.  $\text{Weight}(c)$  indicates the weight of a cue phrase  $c$ . The weight of a cue phrase is calculated by summing up its weight in the knowledge base and the additional weight in the range 0 to 2. If the cue phrase appears at the beginning of a sentence, an additional weight of 2 is added to its original weight retrieved from the knowledge base. If it appears at the

end of a sentence, 1 is added. No additional weight is added if it appears in any other position.

$S_{\text{novel}}(s)$  is the score assigned to the sentence based on the number of novel terms it contains. Weight ( $d$ ) indicates the weight of a novel term  $d$ . The novel terms are detected by searching the terms in to the existing medical vocabulary and English word vocabulary (discussed in section II). If a term is not found in medical vocabulary, we look up into the natural English vocabulary. If it is not found in any of these vocabularies, we treat the term as a novel medical term and assign an empirical weight to the term. But person names, organization names, which are normally not available in the lexicon, create problems to implement the above-mentioned rules for identifying novel terms. Proper names are no doubt important terms, but we did not treat those as novel terms. So, we assign varying weights to the out-of-vocabulary terms. If all letters of an out-of-vocabulary term are capital, we treat the term as highly novel and assign a weight of  $X$  to the term. If all letters are small case and word length  $\geq 10$  characters, we assign a weight of  $(X/2)$  to the term assuming that the term is moderately novel. We found that the word length restriction helped in noise removal while filtering novel terms based on out of vocabulary feature. If only the first character is capital we assign a weight of  $(X/4)$  to the term assuming that the term may be part of proper name or organization name and the novelty of the term is relatively low because we emphasis more on “what says” than “who says”. We set the value of  $X$  to 10 to keep parity with the weight assignment to cue phrases.

$S_{\text{penalty}}(s)$  is basically the negative score assigned to the sentence. Sometimes, the long sentences get more score due to the fact that it contain many words, which may not be of equal importance. Moreover sentences, which are too long indicates the elaboration of a topic. So, we assign a penalty score to a sentence, which is longer than a predefined length.  $L_{\text{min}}(s)$  is the minimum length of sentence worth no penalty. If the length of the sentence is greater than this minimum length it is penalized. We set the minimum length  $L_{\text{min}}$  to 30 words. For example, if the length of the sentence is 35 words it will get a penalty score of 5.

### C. Summary Generation

In this step,  $n$  top ranked sentences are selected to generate the final summary. Value of  $n$  depends on the compression rate. But, the summary produced in this way may contain some redundant information, that is, the some sentences in the summary may entail partially or fully the concept embodied in to other sentences. This restricts the summary to be more informative when the summary length is a restriction. Moreover, a user who is used to just looking at first few sentences representing the same concept will prefer to see something different information, though marginally less relevant. To keep the sentences in the summary sufficiently dissimilar from each other, the diversity based re-ranking method called Maximal Marginal Relevance (MMR) is a well-known measure [16]. This approach offers a ranking parameter that allows the user to slide between relevance to the query and diversity from the sentences seen so far. The

MMR algorithm is most suitable to apply in the query-focused summarization where the summary will be focused toward the user’s query. But in our generic summarization environment where only one summary will be produced for a text document, we used a variant of the MMR algorithm to remove redundancy in the summary. This algorithm works as follows:

- Sort the sentences in decreasing order of their scores
- Select the top ranked sentence first.
- Select the next sentence from the ordered list and include into the summary if this sentence is sufficiently dissimilar to all of the previously selected sentences.
- Continue selecting sentences one by one until the predefined summary length is reached.

The similarity between two sentences is measured using cosine similarity metric. If the cosine similarity between two sentences is greater (less) than a threshold, we say that the sentences are similar (dissimilar).

The cosine similarity [2] is used to measure similarity between two sentences. In our experiment, we set the cosine similarity threshold to 0.7.

Finally, the sentences selected in the above-mentioned manner are reordered using text order (sorted in the order in which they appear in the input texts) [17] to increase the readability of the summary.

## IV. EVALUATION AND EXPERIMENTAL RESULTS

To test our summarization system, we downloaded 50 medical documents from the Internet. Out of 50 medical documents, 20 documents are medical news articles (average size: 2 pages) downloaded from MedLinePlus<sup>1</sup>, 20 documents are medical research articles (average size: 6 pages) and 10 documents are medical reports (average size: 3 pages) downloaded from the website of Indian Journal of Medical Research. For medical news articles no abstract is available. So, for evaluation purpose, we manually generate a reference summary of 150 words for each medical news document. Medical research articles generally come with abstracts. We consider the abstracts available with the research articles as the reference summaries. In effect, for each type of medical document in our test set, we have a reference summary to which the system-generated summary is compared. Before submission of the input document to the summarizer, we remove the author-written abstract available with the document, because our objective is compare the system generated summary with the author written summary to test the performance of the summarization system.

We adopted an automatic summary evaluation metric for comparing system-generated summaries to reference summaries. The method, ROUGE (Recall-Oriented Understudy for Gisting Evaluation)[18], is based on n-gram overlap between the system-produced and reference summaries. ROUGE reports separate scores for 1, 2, 3,

<sup>1</sup> <http://www.nlm.nih.gov/medlineplus/>

and 4-gram matching between the model summaries and the generated summary. Among these different scores, uni-gram-based ROUGE score (ROUGE-1) agrees most with human judgments [18]. When  $n=1$ , the  $n$ -gram is called uni-gram. ROUGE is a recall-based measure and it requires that the summary length be controlled to allow meaningful comparisons, i.e., that is the length of the system generated summary and the reference summary should be the same.

Traditionally, for each document, more than one summary are manually created by a number of human abstractors to test how much the human abstractors agree with each other in producing the summary of the document. In our experiment, we used only one reference summary for evaluation of the performance of our system because a medical research document comes with only one author-written abstract.

To judge the effectiveness of using domain knowledge in the summarization task in medical domain, we set up two experiments.

In experiment 1, we followed summarization approach, which used the well-known features: term frequency, title and position. In this approach, score of a sentence is computed based on the sum of the tf-based weights of the words, its similarity to the title of the document and position of the sentence in the input text.

In experiment2, we implemented the summarization approach, which combines domain specific features (cue phrase and novel term) with the features: term frequency, title and position features used in experiment 1. Equation (1) is used to compute the score of a sentence in this setting.

In fig.2, we have shown an author-written abstract of the medical research article. A system-generated summary of the same article has been shown in fig. 3.

Figure 2. An author-written abstract of a medical research article

Salmonellosis is known to be endemic in India and enteric fever is classically caused by *Salmonella enterica* serotype Typhi. Recently a study on the etiology of enteric fever in north India, New Delhi indicated an increasing trend of *S. Paratyphi A* isolates from 1994 to 1998. An unusually high rate of isolation of *S. Paratyphi A* from the patients suspected to have enteric fever admitted to the Indira Gandhi Medical College & Hospital Nagpur was noticed during 2001-2002. The present study was undertaken to assess the antimicrobial susceptibility pattern of isolates of *S. Paratyphi A* and their phage types. Minimum inhibitory concentration MIC to chloramphenicol was determined by E strip AB Biodisk, Sweden. Chloramphenicol sensitivity documented in literature ranges from 19.7 to 100 percent 72.22 per cent strains in the present study were sensitive to chloramphenicol with MIC values 8 µg/ml. With the substantial increase in drug resistant *S. Paratyphi A* and increased association of this serotype with cases of enteric fever chloramphenicol and ampicillin may not be considered as drugs of choice though they may be indicated in enteric fever caused by *S. Typhi* where more than 90 per cent strains were found to be sensitive. Ciprofloxacin and cephotaxime currently seem to be the drugs of choice in the treatment of enteric fever caused by *S. Paratyphi A*.

Figure 3. A system generated summary of a medical research article

To compare the performance of our system, for each medical document we assume first  $n$  words as the baseline summary where  $n = 150$  for medical news articles and  $n$ =size of the author-written abstract for medical research articles. All model summaries and system-generated summaries are stemmed and stop words are removed while evaluation is done. Table I shows the results after testing our system on a set of medical original research articles. Table II shows the results after testing this system on medical news articles.

TABLE I.  
ROUGE-1 SCORES FOR SUMMARIES ON DATA SET OF MEDICAL RESEARCH ARTICLES

	ROUGE-1 score (average)
Proposed summarization system	0.4013
Summarization system based on TF, Title and position features	0.3889
Baseline	0.3296

TABLE II.  
ROUGE-1 SCORES FOR SUMMARIES ON DATA SET OF MEDICAL NEWS ARTICLES

	ROUGE-1 score (average)
Proposed summarization system	0.4629
Summarization system based on TF, Title and position features	0.4485
Baseline	0.4073

*Salmonella enterica* serotype Paratyphi A has been reported less frequently as a causative agent of enteric fever. Reports on the antimicrobial susceptibility of this pathogen are few and varied. An unusually high occurrence of *S. Paratyphi A* was noted in a tertiary care hospital at Nagpur, Maharashtra during April 2001-September 2002. An effort was made to study the antimicrobial susceptibility pattern and phage types of the isolates. Blood cultures of patients suspected to have enteric fever admitted to the Indira Gandhi Medical College and Hospital, Nagpur were processed by conventional methods. Antimicrobial susceptibility was tested by Kirby-Bauer disc diffusion method and the minimum inhibitory concentration MIC to chloramphenicol was determined. Results: Eighteen 46.15% of 39 *Salmonella* isolates were *S. Paratyphi A* and all were sensitive to ciprofloxacin and cephotaxime. Twelve 66.67% strains were sensitive to ampicillin and 13 72.22 % to chloramphenicol. Two strains 11.11% were resistant to three drugs ampicillin, chloramphenicol and cotrimoxazole simultaneously. The prevalent phage type in the local population was phage type I. The high occurrence of *S. Paratyphi A* found in the present study indicated the emergence of this rare pathogen of enteric fever in the local population. Though some degree of resistance was encountered with ampicillin and chloramphenicol, all the isolates were sensitive to ciprofloxacin, currently a drug of choice for enteric fever. Multidrug resistance was rare.

The results clearly show that incorporation of the domain knowledge in summarizing medical documents improves the summarization performance over the baseline in all of the cases. The proposed system also performs better than the existing system, which summarizes documents using tf, title and position features.

We tested our system on the medical articles of the two types: the medical news articles and medical research articles. We hope that our system will perform well on other types of medical articles (such as commentaries) where abstracts are hardly available.

#### V. CONCLUSION

This paper discusses a text summarization method in medical domain. Most of previous works on text summarization in medical domain used various features used in the news domain. In our work, we used domain specific knowledge in a useful way. We combined the features used in the news domain with the features specific to the medical domain to improve the performance of the summarization task in medical domain. The performance of the system can be further improved by exploring more number of domain specific features and applying learning algorithm for effective feature combination.

#### REFERENCES

- [1] S.D. Afantenos, V. Karkaletsis and P.Stamatopoulos, "Summarization from Medical Documents: A Survey", *Artificial Intelligence in Medicine*, vol. 33, 2005, pp. 157-177.
- [2] I.Mani, "Automatic summarization", *Volume 3 of Natural language processing*, Amsterdam/Philadelphia: John Benjamins Publishing Company, 2001.
- [3] P. B. Baxendale. "Man-made index for technical literature—An experiment". *IBM Journal of Research and Development*, 2(4):354–361, 1958.
- [4] H. P. Edmundson. "New methods in automatic extracting". *Journal of the Association for Computing Machinery*, 16(2):264–285, 1969.
- [5] H. P. Luhn, "The automatic creation of literature abstracts." *IBM Journal of Research Development*, 2(2):159–165, 1958.
- [6] J. Kupiec, J. O. Pedersen, and F. Chen. "A trainable document summarizer". *In Research and Development in Information Retrieval*, pages 68–73, 1995.
- [7] C. D. Paice, and P.A Jones. "The identification of important concepts in highly structured technical papers". *In the proceedings of the 16<sup>th</sup> International Conference on Research and Development in Information Retrieval (SIGIR '93)*, 69-78, 1993.
- [8] H. Jing, and K. McKeown. "The decomposition of human-written summary sentences". *In the Proceedings of SIGIR'99: 22nd International Conference on Research and Development in Information Retrieval*, University of California, Berkeley, August, pp 129–136, 1999.
- [9] H. Jing, "Using hidden Markov modeling to decompose human-written summaries". *Computational Linguistics*, 28(4), 527–543, 2002
- [10] L. Damianos, D.Day, L.Hirschman, R. Kozierok, S. Mardis, T. McEntee, et al. "Real users, real data, real problems: the MiTAP system for monitoring bio Events". *In the Proceedings of the Conference on Unified Science & Technology for Reducing Biological Threats & Countering Terrorism (BTR 2002)*, 2002, PP 167—77.
- [11] D.B.Johnson, Q. Zou, J.D.Dionisio, V.Z.Liu, W.W.Chu. "Modeling medical content for automated summarization". *Ann NYAcad Sci* 2002;980:247—58.
- [12] R.Gaizauskas, P.Herring, M. Oakes, M.Beaulieu, P.Willett, H.Fowkes, et al. "Intelligent access to text: integrating information extraction technology into text browsers". *In Proceedings of the Human Language Technology Conference (HLT 2001)*; 2001. PP. 189—93.
- [13] H. Chen, A. Lally, B. Zhu and M. chau. "HelpfulMed: Intelligent Searching for Medical Information over the Internet", *Journal of American Society for Information Science and Technology (JASIST)*, pages 683-694, volume 54, Issue 7, May, 2003.
- [14] M. Fiszman, T. Rindfleisch, H. Kilicoglu. "Abstraction Summarization for managing the Biomedical Research Literature". *In the Proceedings of the HLT-NAACL Workshop on Computational Lexical Semantics*. 2004.
- [15] P. Chen, R.Verma. "A query-based medical Information summarization system Using Ontology Knowledge". *In the proceedings of the 19<sup>th</sup> IEEE Symposium on Computer based Medical Systems*. 2006
- [16] G. J. Carbonell, and J. Goldstein.. "The use of MMR, diversity-based re-ranking for reordering documents and producing summaries". *In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Melbourne, Australia, pages 335–336, 1998.
- [17] R. Barzilay, Elhadad, and K. McKeown. "Sentence ordering in multi-document summarization". *In Proceedings of the Human Language Technology Conference*. 2001.
- [18] C-Y., Lin, and E. Hovy. "Automatic evaluation of summaries using n-gram co-occurrence". *In Proceedings of Language Technology Conference (HLT-NAACL 2003)*, Edmton, Canada, May 27 - June 1. 2003.