Scientific Research Paper SummarizationOn The Basis Of Research Relevant Term Identification

Sunita R. Patil.
Research Scholar, MPSTME, NMIMS, Parle,Mumbai.

Sunita M.Mahajan.
Principal, ICS, MET, Bandra, Vile Mumbai.

ABSTRACT

This paper presents few additions to our existing system for summarizing multiple scientific research papers published at various recognized journals, conferences and workshops informing latest research developments.This modified system is more useful as compare to previous oneguiding many researchers or research scholars looking for innovative contributions in a specific field of research. The similarity in contents and repeated relevant information from multiple domain specific scientific articles are reduced and optimized by novel research term analysis method.

Two new innovative categories of Research Relevant Novelty [RRN] terms, uniqueness and difference from previous ideas (like and contrast) and research continuations of earlier/existing work (continuation/novel) are added to existing system. This modified version minimizes information overload problem present in this online era by providing most condensed, accurate, optimized and relevant contents from multiple scientific papers through the existing categoriessuch as research purpose (aim), approach & methodology used (method), and results & discussions (outcome).This up gradation results in a effective and efficient strategy for minimizing scholars efforts in reading all scientific papers completely to get desired information.

Keywords:

summarization, optimization, research relevant novelty, scientific research papers.

1. INTRODUCTION

In our previous innovation we developed a strategy for summarizing domain specific scientific research articles. These articles are published at most of the international journals, conferences or workshops presenting latest research developments in the domain field. Whenever a new research scholar or a researcher wants to refer the existing works, methods and technologies for his/her reference, they search for the field specific published technical research papers. The problems associated with these multiple papers are similarity in contents and the repeated relevant information. Thus reading them all completely one by one is time-consuming, unnecessary and impossible.

Our previous multi-document summarization system [S.Patil., S. Mahajan, 2011] showed very good results to solve above problems. This time we have added two more categories to completely fulfill scholar expectations. Thus this up gradation expands our research work for getting short, condensed and accurate information from multiple published papers informing previous and latest research developments in the field of interest.

Multi-document summarization is the process of dealing with a large amount of information present in multiple related source documents by comprises only the essential material or main ideas in a document in less space [Mani, 1999]. We are using the same concept with innovative ideas for summarizing multiple domain specific research papers.

Our system works on first two sections of research papers which are *abstract* and *introduction* sections only. Since every journals or conferences publishes author's work with the format consisting these two sections mandatory. We analyzed that the sentences from these two sections also appears else wherein the body of the document either as a close variant or in identical form.The '*abstract*' section of the paper focuses on the purpose of research; where-in the statement of the problem(s) or research issue(s) are addressed with the research methods used (experimental research, case studies, questionnaires, etc.); the results and/or findings of the research; and the main conclusions and recommendations, therefore briefly summarizes the workdone.

Table I: Upgraded Research Categories

Goal	Sentences representing the purpose or aim or major innovative idea of research under study for current paper;
Method	Sentences representing the methods or approaches or ways used for the goal achievement;
Contrast & Like	Sentences claiming authors own work contrast with others/ earlier work; sentences showing limitations in others/ earlier work; direct comparing with others/ earlier work; research work of this kind never done before; sentences presenting similarity with others / earlier work;
Continuation	Sentences describing research continuation of earlier/existing work;
Outcome	Sentences relating to result, conclusion, outcome; Sentences showing end product; Sentences stating evaluation of implementation;

The next is '*introduction*' section which provides general description about the importance of the topic and its history in the field; establishes the context of the work being reported; this is accomplished by discussing the relevantprimary research literature (with citations) and summarizing current understanding of the problem under investigation; State the purpose of the work in the theory, question, or problem investigated; and, briefly explain rationale behind each step; the approach and, whenever possible, the possible outcomes the study can be revealed. It also contains information that allows the reader to fully understand the paper topic, the topic's relevance, and the paper's thesis before proceeding to more in-depth examination.

The main focus of our research is to present the summary for multiple related domains specific research articles under previously identified three categories: the innovative scientific ideas (research aim), approaches & methodologies (technique), results & discussion (outcome) and newly added two more categories: uniqueness and difference from previous ideas (like and contrast) and research continuations of earlier/existing work (continuation/novel) as shown in table I.

Two newly added categories improves scholars understanding about the research article in terms of authors similar kind of previous work continuation and contrast with others/ earlier work done in the same field of study.

2. LITERATURE SURVEY

Summarization work had started long back with the need of articlesummaries which produces identification of recurrent descriptions of the same events**.** Until recently, generic summaries were more popular, but with the prevalence of full-text searching and personalized information filtering, user-focused summaries are gaining importance. The pioneers of automated summaries [Luhn, 1958; Baxendale, 1958; Edmundson, 1969;Barzilay, McKeown, and Elhadad (1999] used different approaches such as frequency-based, sentence position, or rhetorical clue features in their study.They also predicted that their approaches can be used for the automatic creation of scientific summaries. Further**[**Schwartz and Hearst, 2006;Mei and Zhai, 2008 and Qazvinian and Radev, 2008] used citation information in creating summaries for scientific articles.In recent work [Teufel 1999; Teufel and Moens, 2002**;**Teufel et al.2009andAngrosh et al. 2010] defined and studied argumentative zoning of texts. They studied the structure of an entire article containing background knowledge (BACKGROUND zone)andothers work (OTHER and BASIS zones).

Automated creation of technical surveys from a research topic [Mohammad et al., 2009] uses standard generic multi-document summarization algorithms. They showed that citation information was effective in the summarygeneration process. Recently [Nakov et al., 2004, Nallapati et al., 2008] also showed that the citing sentences in other papers can give a useful description of a target work.Swale's, 1990 moves described the rhetorical status of a text segment with respect to the overall message of the document with the CARS model showing how patterns of these moves can be used to describe the rhetorical structure of introduction sections of physics articles. Summarizing sociology dissertation abstracts [ShiyanOu, Christopher S.G. Khoo, 2007]using semantic-level research variables, their relationships, taxonomy construction reports the similar type of work.

As stated in our previous system[S.Patil., S. Mahajan, 2011], we focuses on identifying 'Research Relevant Novelty' [RRN] terms present in sentences concentrating research purpose/goal/aim, research methodologies/approached/techniques used and the outcome/result of the paper. This time we identified RRN terms representing uniqueness and difference from previous ideas (like and contrast) and research continuations of earlier/existing work (continuation/novel) also. Adding sentences concentrating on these two categories expands usefulness of summary. The use of 'Maximal Marginal Relevance' (MMR) strategy [Goldstain et. al. and Carbonell, 1998] will be the same for summarizing these research papers as redundancy and optimization are the main concerns of our study.

3. SYSTEMIMPLEMENTATION

Our system comprised of four main steps: pre-processing, information selection and categorization and summary generation[Mani, 1999].The novel approaches proposed in this research would mainly be in the *sentence selection and categorization.* Previous summarization approaches selects important information using rhetorical relations between sentences and presents the source paper's objectives and contributions [S. Teufel and M. Moens, 2002] another approach extracts research concepts and their relationships in the text [S. Ou, C.S. Khoo, and D.H. Goh,2008]few more approaches integrates relevance of the subject by measuring correlation between the topic and the source content [Mani, 1999], the summary topic [D. Zajic, B. J. Dorr, J. Lin, R. Schwartz, 2007] or the user query [T. He, W. Shao, H. Xiao, and P. Hu,2007].In one more approach comparing source sentences and measuring semantic similarity, unique information across the source documents can be identified [J.D. Schlesinger, D.P. O'leary, and J. M. Conroy, 2010] or the other approach measures the relative information gain ratios of information with respect to the adjacent text [R. Feldman and J. Sanger,2006]. The system algorithm basically consists of seven steps as explained below.

Step1.Preprocessing:

Preprocessing step consists of activities such as: Segmenting research papers into 'A*bstract'* and *'Introduction '*sections only. Identifying and removing formulas, tables, figures, eventual LATEX mark ups if any and citations from text files. Detecting sentence boundaries and in turn splitting sentences into words.Tokenization is done through stemming [32], removal of stop/noisy words and punctuation from indexed data. Thus text is cleaned for automatic processing.

Step2.RRNTerm Identification:

This is the upgraded innovation where we are identifying sentences relevant to the query using any string similarity algorithm like cosine similarity matrix [Salton and McGill, 1983] with a threshold below which no sentences will be selected to be added into the summary. We identified individual words or phrases called 'Research Relevant Novelty' (RRN) terms for each category as shown in table I,reflecting their significance in the text and having relative complete meaningwhich isnothing but the research concentrated 'novelty' presentations from papers. We divided all sentences from 'A*bstract'* and *'Introduction'* sections into previously identified three categories such as *research goal, research methods*used, and*outcome*[S.Patil., S. Mahajan, 2011] plus newly identified two more categories such as *contrast & like* and *continuation* of the paperwhich are words or phrases representing theses categories as term t*ypes*.

Step3. Similarity Measure and Sentence Clustering:

All the sentences starting with or containing RRN terms of each category are extracted. Sentences containing similar category are clustered together. Using 'Maximal Marginal Relevance' (MMR) metric to find similarity between multiple sentences[Goldstain et. al. and Carbonell, 1998].MMR metric is defined as,

$$MMR(P,C,Q,R,S) = Arg\max_{P_{ij} \in R \setminus S}[\lambda * Sim_1(P_{ij},Q,C_{ij}) - (1-\lambda) * \max_{P_{nm} \in S}(Sim_2(P_{ij},P_{nm},C,S))] \quad (1)$$

Where *D*representsdocument collection, P_{ij} is the sentence j from documentD_i, C_{ij} is the subset of clusters of C that contains sentences P_{ij}, $Q$is the query/topic specification, $w_i$is the weights for the terms to be optimized, Wstands for the word / term in the sentence P_{ij}, $type$is the particular type of word / term, R = IR (D, P, Q, Θ) stands for ranked list of sentences from documents retrieved by IR system, where Θ is the relevance threshold, below which it will not retrieve sentences , S is the subset of sentences in R already selected and $R\backslash S$ is the set difference i.e. set of sentences in R not yet selected. To compute standard relevance ranked list plus some additional scoring factors set λ =1 and to compute maximal diversity ranking among the documents in R set λ =0.Sim_1, similarity metric is calculated for relevance ranking as sum of *cosine similarity* metric of the sentence and query , *coverage* score for the sentence specifying whether the sentence is *in one or more clusters* and the *size* of the cluster and *information content* of the sentence by taking into account the *RRN terms*.Sim_2, similarity metric for anti-redundancy is calculated as the cosine *similarity* metric of sentence and previously selected sentence, *penalize sentences* that are part of clusters from which other sentences have already been chosen and *penalize documents* from which *sentences* have *already been selected*.

Step4. Sentence Selection and Sentence Scoring:

Selecting representative sentences from the clusters is a key problem. In general, there are two kinds of search strategies [ShiyanOu, Khoo and D .H. Goh., 2007]. The local strategy tries to find a representative sentence for each cluster based on the information configuration of the cluster itself, while the global strategy tries to find the representative sentence based on the overall performance of the whole summary. For each sentence cluster, select one sentence to represent the category denoted by the cluster. Each sentence denote the Document ID number, Sentence Number to be needed by final summary. Assign weights to the sentence based on the terms included in it. As per 'local search strategy' select the representative sentence based on the clusters category themselves.

i. Local Search Strategy

a. Centroid Sentence :

Centroid sentence is selected by two steps.

i. First, the centroid vector of the cluster is calculated.
ii. Second, the sentence, which has the smallest distance with the centroid vector, is selected using cosine distance.

b. Calculate Tem Weight **(W_{ij})**:

Each sentence P_{ij} is represented as the weights of terms,$Wij. Wij = (w_{i1}, w_{i2}, \ldots, w_{iN}),\ i = 1,2, \ldots M$, where M is the number of sentences and N is the number of total terms in collection, w_{ij} i.e. the weight of the jth term in the ith sentence(term weight) is calculated as the normalized term frequency in the sentenceP_{ij}.

$$W_{ij} = TF(term_{ij}) / \sqrt{\sum_{t=1}^{n} \mathrm{TF}^2(\mathrm{term_{ij}})} \quad (2)$$

Where $TF\,(term_{ij})$ denotes the occurrence number of the jth term in the ith sentence i.e the measure of how often a term is found in a collection of documents. $TF\,(term_{ij})$is combined with Inverse Document Frequency (*IDF)* as a means of determining which documents are most relevant to user query.

ii. Global Search Strategy

Global search strategy, selects a sentence according to its contribution to the performance of the whole summary. For this a global criterion is needed to measure the summary. The criterion is defined as follows:

$$w_{summary} = \frac{\sum_{t\epsilon summary} \log\left(1+TF\left(term_{ij}\right)^{D}\right)*\log\left(1+l_t\right)}{\log\left(1+l_{summary}\right)} \quad (3)$$

Where$t$,is the term in the summary, $TF\left(term_{ij}\right)^D$is term frequency in document collection, $l_t$is the term length. Intuitively, the criterion reflects the global term density of a summary. In general, we expect the summary to contain more terms, more longer terms, and as short as possible in each selecting step.

In above both cases sentence subsuming is done i.e. if the information content of sentence '**a'**contained within sentence '**b'**, then '**a'**becomes informationally redundant and the content of '**b'**are to subsume that of '**a'** with the additional contents.

Once the sentences are selected, three features used for scoring them to be included into summary:

Let $S_{i,k}$ denote the i[th] sentence in the document D_k belonging to$D$(collection of docs), then we define three features for scoring the sentences:

i. ***Centroid value***

The centroid value for sentence $P_{i,k}$ is defined as the normalized sum of the centroid components.

$$Ci,k \;= \sum_{w\in S_{i,k}} (\,[TF(term_i)\; IDF(term_i)]/ |D|\,) \quad (4)$$

ii. ***Positional Value***

For every sentence $S_{i,k}$suppose it is coming from documentD_k, where length (n) is the number of sentences in D_k . The positional value for this sentence is computed as:

$$Pi,k = ((n - i + 1)/n) * C \quad (5)$$

iii. ***Sentence Overlap Value***

The overlap value is computed as the inner product of the sentence vectors for the current sentence i and the first sentence of the document. The sentence vectors are the n dimensional representations of the words in each sentence, whereby the value at position $i$of a sentence vector indicates the number of occurrences of that word in the sentence.

$$Fi,k = S_{i,k}\,.\,S_{1,k} \quad (6)$$

The raw score for each sentence will be addition of centroid, positional and sentence overlap value with sentence weight.

$$SCORE\,(Si) = wcCi,k + wpPi,k + wfFi,k \quad (7)$$

Where*i (1<= i<= n)* is the sentence number within the cluster.

wc, wp, wf are weights of the features.

Step5. Sentence Comparison Scoring:

The raw score of each sentence is calculated in above step and is compared between multiple sentences from each cluster. Score comparison can be arranged in ascending or descending order for multiple sentences from multiple papers or scored sentences from individual paper summaries. The highest scored sentences among all categories are selected summarization.

Step6.Multi-Paper Summary Generation:

Using any one summary cohesion criteria i.e. the ability to combine research oriented 'Relevant Novel' sentence text in a useful for the user or reordering all sentences in each cluster, sentences are selected with higher score and rebuilt into multiple paper summaries. This cohesion can be d*ocument ordering, rank ordering, topic-cohesion and occurrence ordering.*For final multi-paper summarization, the elected sentences are stored till the desired percentage of summarization met. In order to generate desired percentage of summary, a threshold is set as: ((Total sentences of first document) + (Total sentences of second document) + (Total sentences of third document) +…+ (Total sentences of n^{th} document)) × (Desired percentage of summary).

4. EVALUATION

The evaluation is measuring and comparing system performance with existing summaries. One must also evaluate the qualitative and quantitative properties of the summaries to determine its exact usefulness. An ideal summary must possess at least few fundamental properties such as:

- Compression Ratio-The ability to find how much the summary is shorter than the original.
- Retention Ratio-The ability to find how much of the central information is retained i.e. identifying coherence and readability of the text.
- Accuracy - The ability to find and extract the desired important information across the documents.
- Conciseness - The ability to minimize redundancy between candidate sentences and,
- Optimization- The ability to enhance effectiveness of summary, to make it functional at its best or most useful.

Summary evaluation methods can be divided into two fundamental categories: intrinsic and extrinsic [Spark-Jones and Galliers, 1995; Mani and Maybury, 1999].

- Intrinsic evaluation measures the quality of summaries directly (e.g., by comparing them to ideal summaries).
- Extrinsic methods i.e. task-based measures how well the summaries help in performing a particular task (e.g., clustering).

There are two general types of summaries used for comparison with the automatic summaries being evaluated.

- First, *gold standard summaries* (or *target summaries*) can be author summaries, professional summaries or summaries produced specifically for the evaluation.
- Second, *baseline summaries* are generally produced by extracting random sentences from source texts or produced by another system.

This automated summary is compared against an "ideal" summary. To construct the ideal summary, a group of human experts were asked to extract relevant sentences. The sentences chosen by majority of humans are included in the ideal summary. We should note that [Jing et al., 1998] the cut-off summary length can affect results significantly, and the assumption of a single "ideal" summary may be problematic.

4.1 Corpus:

The data set contains retrieved research papers from IEEE Explore on 'Information Retrieval' domain and relevant to topic specific query on any term research scholar wants know detailed information about.

4.2 Human agreement (A):

Three sets of gold standard data were manually created from *Abstracts* and *Introduction* sections respectively:

i. Human annotators (knowledgeable in scientific article review generation) were asked to identify important sentences of a single paper worth included in summary.
ii. Human annotators verified relevance creation.
iii. Determining how well the different automatically generated summaries performed against these gold standards.
The annotators' agreement measured to what extent each annotator satisfies the utility of the other annotator by picking the right sentences.

5. EXPERIMENTAL RESULTS

We used average precision which is a widely used measure to evaluate information retrieval systems. It is computed by measuring both *recall* and *precision* at various points. *Precision* is the ratio of relevant retrieved documents to retrieved documents while *recall* is the ratio of relevant retrieved documents to relevant documents.*F*-measure [van Rijsbergen, 1979] is a convenient way for reporting *precision (P)* and *recall (R)* in one value. For evaluating performance average across all five categories and overall performance, the scores are first determined by computing performance measures per category then averaging these to compute the global means. Secondly in particularperformance, scores are determined by computing the total average scores for all categories then using these totals to compute the performance measures.

$$F1 = \frac{2*precision\ *recall}{(precision\ +recall\ )} \quad (8)$$

Table II below show the F-measure score computed for automated summary against human summary for each category identified with the help of RRN terms.

Table II. *F*-measure Score for Human generated and Automated Summaries.

RRN Categories	*F*-measure	
	(Human Summary)	(Automatic Summary)
Goal	62%	51%
Method	51%	28%
Contrast & Like	79%	60%
Continuation	92%	86%
Outcome	71%	45%

The result shows that the automatic system obtains substantial improvement over the human in terms of precision and recall over the categories Goal, Method, Contrast & Like, Continuation and Outcome.

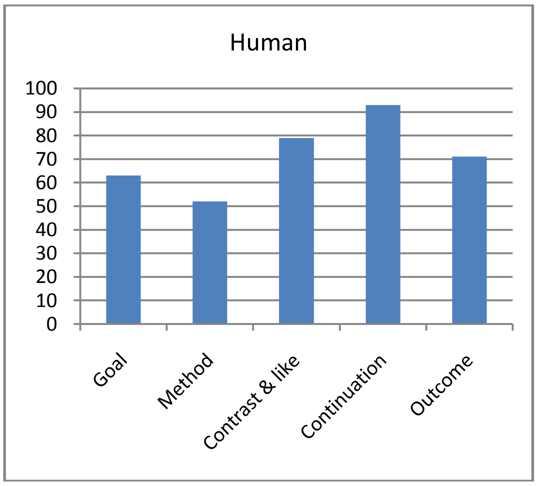

Figure 1. Performance per research category: *F*-measure-Human Summary

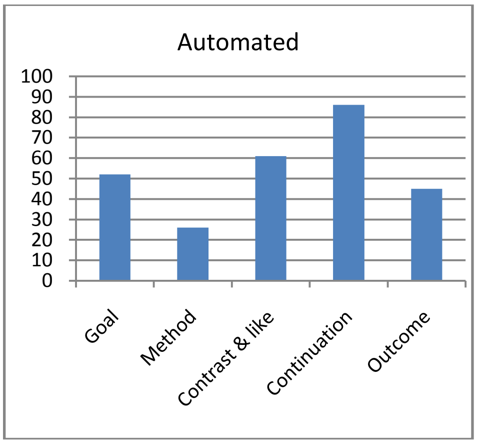

Figure 2. Performance per research category:*F*-measure-Automatic Summary.

So far automatic summarization has not yet reached the quality possible with manual summarization, where a human understand the text and writes a completely new shorter text with new lexical and syntactic choices. However, automatic summarization is persistent, reliable and always available.

6. CONCLUSION

Summarizing multiple research papers using 'Research Relevant Novel' term identification for various research categories by analyzing few important queries such as, what is the goal/purpose of this paper? What are the techniques/methodologies used for implementation? What are the similarities /contrast between multiple relevant author works? Whether the paper is a continuation of previous innovation?Andwhat are the outcomes /results presented in these multiple relevant papers on same user query? Thus this system is an aid for research scholars for getting short, condensed, accurate, explicit, optimized and most relevant summarized information from domain-specific topic-based multiple research papers. The system informs earlier and latest research developments, progress, challenges and future scope in the particular field of study using innovative 'Research Relevant Novelty' (RRN) term analysis i.e. the main contribution presented by authors in the research papers through various categories such as *research goal*, *research methods* used, *contrast & like,continuation* and *outcome*.

This upgraded system provides starting material for research scholar for further innovation specifying current research methods/techniques/approaches used, compared with others.This work also addressed redundancy problem using 'Maximal Marginal Relevance' and introduced 'Research Relevant Novelty' term identification for simplicity and efficient summary generation.

REFERENCES:

[1] Angrosh, M. A., Stephen Cranefield, and Nigel Stanger (2010), Context identification of sentences in related work sections using a conditional random field: towards intelligent digital libraries. In JCDL'10: Proceedings of the 10th annual joint conferenceon Digital libraries, pages 293–302, ACM.

[2] Barzilay, R., McKeown, K., &Elhadad, M. (1999). Information fusion in the context of multi-document summarization. In *Proc. of the 37th Annual Meeting of the Assoc. of Computational Linguistics*.

[3] Baxendale, P (1958), Machine-made index for technical literature - an experiment. IBM Journal of Research Development, 2(4):354{361.

[4] D. Zajic, B. J. Dorr, J. Lin, R. Schwartz. (2007) "Multi-candidate reduction: Sentence compression as a tool for document summarization tasks". *Inf. Process. Manage.*, vol. 43, no. 6, pp. 1549-1570.

[5] Dragomir R. Radev, Hongyan Jing, MałgorzataSty_s, Daniel Tam, (2004) Centroid-based summarization of multiple documents, Information Processing and Management, 919–938.

[6] Edmundson, H. P (1969), New methods in automatic extracting. Journal of the ACM, 16(2):264-285.

[7] Goldstainet. al. and Carbonell, (1998), The Use of MMR, Diversity Based Re-ranking for reordering documents and producing summaries. In proceedings of SIGIR 1998, Melbourne, Australia .p335-336.

[8] J.D. Schlesinger, D.P. O'leary, and J. M. Conroy, (2010), "Arabic/English Multi-document Summarization with

CLASSY – The Past and the Future". *Lecture Notes in Computer Science*, vol. 4919, pp. 568-581.

[9] Jing, Hongyan, Regina Barzilay, Kathleen R. McKeown, and Michael Elhadad(1998), Summarization evaluation methods: Experiments and analysis. Working Notes ofthe AAAI Spring Symposium on IntelligentText Summarization, pages 60–68.

[10] Lin, Chin-Yew and Eduard Hovy (2003), Automatic Evaluation of Summaries Using N-gram Co-Occurrence Statistics, Proceedings of the Human Technology Conference (HLTNAACL-2003), Edmonton, Canada.

[11] Luhn. H.P. (1958), "The Automatic Creation of Literature Abstracts". IBM Journal of Research and Development, Vol. 2, No. 2, pp. 159-165.

[12] Mani and M. T. Maybury, (1999) "Introduction". In *Advances in Automatic Text Summarization*, I. Mani and M.T. Maybury, Eds. Cambridge, MA: MIT Press.

[13] Inderjeet MANI,Summarization Evaluation: An Overview,The MITRE Corporation, W640.

[14] McKeown, Kathleen R. and Radev (1995), Generating summaries of multiple News Articles. In ACM-RDIR, SIGIR'95, WA.

[15] Mei, Qiaozhu and ChengXiangZhai(2008), Generating impact-based summaries for scientific literature. In *Proceedings of ACL '08*, pages 816–824.

[16] Mohammad, Saif, Bonnie Dorr,Melissa Egan, Ahmed Hassan, PradeepMuthukrishan, VahedQazvinian, DragomirRadev, and David Zajic (2009), Using citations to generate surveys of scientific paradigms.In*NAACL 2009*, pages 584–592.

[17] Nallapati, R. M., A. Ahmed, E. P. Xing, and W. W. Cohen (2008), Joint latent topic models for text and citations. In Proceeding of the 14th ACM SIGKDDInternational Conference on Knowledge Discoveryin Data and Data Mining, pages 542–550.

[18] Nakov, A. S. Schwartz, and M. Hearst (2004),Citances: Citation sentences for semantic analysis of bioscience text. In*Proceedings of the SIGIR'04 workshop on Search and Discovery in Bioinformatics.*

[19] S. Teufel and M. Moens (2002), "Summarizing scientific articles: experiments with relevance and rhetorical status," *Comput. Linguist.* vol. 28, no. 4, pp. 409-445.

[20] Salton, G. and Buckley, C. (1987), Term weighting approaches in automatic text retrieval. Information Processing and Management, 24(5), 513-523.

[21] S. Ou, C.S. Khoo, and D.H. Goh (2008), "Design and development of a concept-based multi-document summarization system for research abstracts," *J. Inf. Sci.*, vol. 34, no. 3, pp. 308-326.

[22] S. Ou, Khoo, C. &Goh, D(2007), Automatic multi-document summarization of research abstracts: design and user evaluation. *Journal of the American Society for Information Science and Technology (JASIST)*, 58(10). Pp. 1-17.

[23] Schwartz and Marti Hearst (2006), Summarizing Key Concepts using Citation Sentences *Proceedings of the BioNLP Workshop on Linking Natural Language Processing and Biology at HLT-NAACL 06*, pages 134–135,New York City, Association for Computational Linguistics

[24] Spark-Jones, K. and J. R. Galliers (1995), Evaluating Natural Language Processing Systems: An Analysis and Review. Number 1083, Lecture Notes in Artificial Intelligence, Springer.

[25] SunitaPatil,SunitaMahajan (2011), Summarizing Domain Specific MultipleReseach Papers Using Research Relevant Novelty selected at ICDM'11, Ghaziabad, India.

[26] Teufel, Simone, AdvaithSiddharthan, and Colin Batchelor (2009), Towards domain-independent argumentative zoning: Evidence from chemistry and computational linguistics. In Proceedings of Conference on Empirical Methods in NaturalLanguage Processing, pages 1493–1502, Singapore, ACL.

[27] Teufel Simone and Marc Moens (2000), What's yours and what's mine: Determining intellectual attribution in scientific text. In Proceedings of the JointSIGDAT Conference on Empirical Methods inNatural Language Processing and Very LargeCorpora.

[28] Teufel, (1999), 'Argumentative Zoning: Information Extraction from Scientific Text'.

[29] T. He, W. Shao, H. Xiao, and P. Hu (2007), "The Implementation of a Query-Directed Multi-document Summarization System," in *Proceedings of the Sixth international Conference on Advanced Language Processing and Web information Technology,* pp. 105-110.

[30] R. Feldman and J. Sanger (2006)*The Text Mining Handbook: Advanced Approaches to Analyzing Unstructured Data*. Cambridge, England: Cambridge University Press.

[31] Qazvinian, Vahed and Dragomir R. Radev (2008), Scientific paper summarization using citation summary networks. In Proceedings of International Conferenceon Computational Linguistics (COLING), pages 689–696, Manchester, UK.

[32] vanRijsbergen, CornelisJoost (1979), Information Retrieval. Butterworth, London, second edition.