

CLUSTER PRIORITY BASED SENTENCE RANKING FOR EFFICIENT EXTRACTIVE TEXT SUMMARIES

Yogesh Kumar Meena¹ and Dinesh Gopalani²

Department of Computer Science & Engineering, MNIT, Jaipur, India

ABSTRACT

This paper presents a cluster priority ranking based approach for extractive automatic text summarization that aggregates different cluster ranks for final sentence scoring. This approach does not require any learning, feature weighting and semantic processing. Surface level features combinations are used for individual cluster scoring. Proposed approach produces quality summaries without using title feature. Experimental results on DUC 2002 dataset proves robustness of proposed approach as compared to other surface level approaches using ROUGE evaluation matrices.

KEYWORDS

Sentence Ranking, Extractive, Automatic Text Summarization, Information Retrieval, Summary Generation.

1. INTRODUCTION

With the increasing load of textual information over the WWW, it becomes time consuming for the users to find the desired information need. They have to read whole the document to get a glimpse of its actual theme. Automatic text summarization solves the problem by generating summaries that could be utilized as a condensed replica of a document or a set of documents. Automatic Text Summarization (ATS) is defined as the process of condensing the source text document while retaining its main information contents using a computer machine. Automatic text summarization is classified mainly as extractive and abstractive automatic text summarization [1,2,3]. Research in the field of text summarization started from 1950's and as of now no system is available that can generate summaries as like professionals or humans. Most of the researchers focused on generic extractive summarization in their research contributions. Generic summaries basically represent overall information contents in condensed manner. The basic process flow of generic Extractive text summarization is shown in Figure 1. Preprocessing is the first Step in which sentences are segmented. In general '.' and '?' symbols are used as end of sentence markers. Stop words are removed thereafter. Stemming is performed using standard method like Porter's Stemming Algorithm.



Figure 1. Generic Extractive Summarization Process

After Preprocessing the sentences are scored using scoring functions. These sentences after scoring are selected in the same order as they present in the original document in the final summary generation phase. Besides generic summarization, update summaries, query focused summaries, sentimental summaries etc. can also be extracted from the text, but it depends on the purpose of summarization process. There are few efforts made in abstractive automatic text

summarization as well, but due the requirement of large domain knowledge, it is not preferred. Another reason is that information fusion, which is the most difficult part of abstractive text summarization. Statistical methods for extractive ATS do not use much domain knowledge but instead they work on the available information content on the source document.

Most of the approaches proposed by researchers required to use some weight for features to calculate final sentence scoring. This makes it difficult to take advantage of these features uniformly for all documents. Our approach does not use feature weights, instead of this our approach does numerical processing and use cluster priority based ranking that alleviates the issue of feature weighting. Our approach does not used title feature as it is not necessary that all text documents are provided with their tile or heading. Proposed approach initially scores sentences using linear sum of different optimal features. Thereafter computes sum of inverse scores of individual feature combination clusters. Our approach using ROUGE evaluation matrices performs better as compared to other existing feature based approaches.

Rest of the paper is organized as; in section 2 we discuss related work in the area of extractive ATS. In section 3 we discuss features used. Proposed method is discussed in Section 4. Section 5 discusses results and performance evaluation. In section 6 we conclude the paper.

2. RELATED WORK

Research in the area of automatic text summarization started in 1950s but still we are lacking with efficient methods that can generate summaries like humans professionals. Luhn [4] in 1958 proposed the method that utilized term frequency to score the sentences. This idea was criticised later that only term frequency cannot be a good idea to score the sentences. Later Baxendale [5] in 1969 proposed sentence location as a scoring criterion along with term frequency to score the sentences that improved the results. To further improve the efficiency two more features were proposed by Edmundsun [6] in 1969. Title similarity and Cue word features are also included for sentence scoring along with earlier two for efficiency improvement. Rush et al [7] in 1971 used sentence rejection methodology using different set of rules that was later used in 1975 for generation of chemical abstracts [15]. In 2001 MEAD [9] was developed using features such as TF/IDF, Sentence Location, Cue words and Longest Common Subsequence's (LCS) for sentence scoring. Later researchers used additional features and used different techniques [8,10,11,12,14,16,21] for sentence scoring such as gain[21], modified LSA[22]. Various review articles [1,2,3,17,18,19] are also published to analyze the performance of these features. All above researchers scored sentences on the basis of some features. Most of the researchers gave equal weight to the all features included for sentence scoring.

3. FEATURES USED

Sentence level scoring methods basically use a set of features (mainly surface level) to score the sentences. These features are extracted from pre-processed source document. These features capture the important information content available in the input source text document. Features used for our proposed approaches are described as follows:

3.1 Term Frequency (F1)

It is the frequency of the term in the whole text document. According to this feature the most frequent words in document are the most important words (other than stop words) and they convey maximum information.

3.2 Gain (F2)

Words with medium (Stop words are higher frequency words) frequency are most important, gain used this idea effectively for sentence scoring.

3.3 Word Co-occurrence (F3)

Thematic words (most frequent) if they co-occur in the sentences then higher weight is given to the respective sentences.

3.4 Sentence Location (F4)

It assigns a score to each sentence as per its location in the text document. First sentence always got higher score in this feature.

3.5 Named Entity (F5)

Presence of named in a sentence, suggests that it contains more information and should be scored higher.

3.6 Sentence Centrality (F6)

The sentences which are more similar (Cosine, Jaccard etc.) to other sentences are the most important ones. The similarity is measured in terms of terms overlap between the sentences.

3.7 Text Rank (F7)

It is based on one of the most popular ranking used for web link analysis and assigns the scores to sentences considering them as nodes for link analysis.

3.7 LSA (F8)

Modified latent semantic scores procedure is used for scoring sentences using this feature as proposed by Wang and Maches[31].

3.9 Bushy Path (F9)

According to This feature the sentence that have maximum sentences related (with specific threshold value) to it, is the most important sentence.

4. SUMMARIZATION PROCESS

Our Proposed approach used optimal features for sentence scoring. These optimal features identify important sentences in the source text document. Our proposed summarization process completes in steps as shown in Figure 2. In first step, pre-processing is applied on input source document. Operations such as sentence segmentation, stop word removal, stemming and case folding are applied. We then use Stanford NLP tools for POS tagging and to find named entities in the text. After pre-processing we use feature clusters for scoring process as given in equations 1-6. We used optimal features like term frequency, gain, sentence location, LSA, named entity bushy path, textrank , sentence centrality and word co-occurrence to create combinations. Using these combinations we have created feature combination clusters. As given in equation 1-4, we

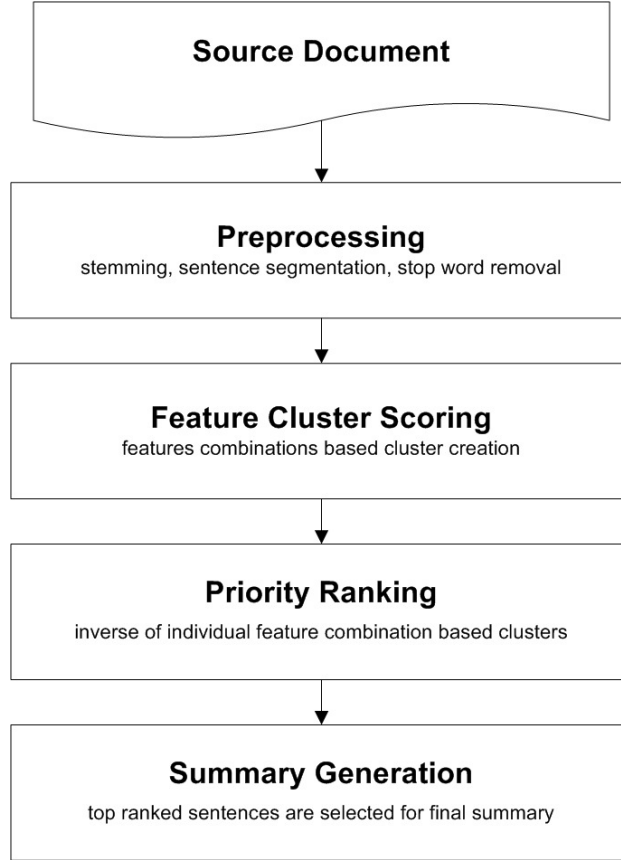


Figure 2. Flow Chart for Proposed Summarization Process

created four clusters using combinations of different features that optimize the quality of summary. Further, we used their inverse scores as proposed score1 and proposed score 2 as given in equation 5 and equation 6. After scoring the sentences using proposed score1 and proposed score 2, we ranked each of the sentences. Top ranked sentences are selected as the summary sentences in final summary generation step.

$$score_1(s) = F1(s) + F2(s) + F4(s) + F5(s) + F6(s) + F8(s) + F9(s) \quad (1)$$

$$score_2(s) = F1(s) + F2(s) + F5(s) + F6(s) + F7(s) + F9(s) \quad (2)$$

$$score_3(s) = F1(s) + F5(s) + F8(s) + F9(s) \quad (3)$$

$$score_4(s) = F1(s) + F3(s) + F5(s) + F9(s) \quad (4)$$

$$proposedscore_1(s) = \frac{1}{score_1(s)} + \frac{1}{score_2(s)} + \frac{1}{score_3(s)} + \frac{1}{score_4(s)} \quad (5)$$

$$proposedscore_2(s) = \frac{1}{score_1(s)} + \frac{1}{score_2(s)} \quad (6)$$

Our approach does not require any kind of semantic processing. Title feature is also not included as sometimes title of source document is unavailable.

5. RESULTS AND ANALYSIS

We experimented our proposed scoring methods on DUC 2002 dataset. We used a subset of 10 documents for our experiments. Each document is given with a reference summary (gold summary), which we used to compare with our system generated summary. We limited the compression rate to 100 words. While generating summary, if total words in summary are exceeded the value 100, sentence inclusion is stopped. We used Recall-Oriented Understudy for Gisting Evaluation (ROUGE) [20] matrices to evaluate quality of generated summaries.

$$ROUGE - N = \frac{\sum_{S \in GoldSummary} \sum_{gram_n \in S} count_{match}(gram_n)}{\sum_{S \in GoldSummary} \sum_{gram_n \in S} count(gram_n)} \quad (7)$$

Here ROUGE-N check n-gram co-occurrence in between system generated and gold summaries as given in equation 7. We experimented on ROUGE-1 and ROUGE-2. We used ROUGE-L which finds longest common subsequence. This identifies longest co-occurring in sequence of n-grams automatically. In ROUGE-W, weighted LCS is used that favours consecutive longest common subsequences. We used term frequency, linear sum of all features described in section 3, MEAD, MS word and modified LSA for comparative analysis with our proposed scoring techniques. Our approach is not using title feature and semantic processing.

The results of ROUGE-1 recall of different techniques and proposed scoring methods are presented in Table-1. Comparing with individual documents our proposed method2 performs better for document 3,5,6,7,8,10 as compared to TF, Linear Sum, MEAD, MS Word and LSA. However, for documents 1,2 rest of the other methods result in the similar range and higher as compared to proposed methods. Proposed method2 takes the advantage of optimal features with their combinations. With these feature combination clusters using inverse ranking process selects informative sentences that ultimately improves the quality of produced summaries.

Table 1. ROUGE-1 recall scores of different techniques and proposed approaches.

Document Number	TF	Linear Sum	MEAD	MS Word	LSA	Proposed Method1	Proposed Method2
Doc1	0.51961	0.50980	0.50980	0.50980	0.50000	0.50980	0.50980
Doc2	0.42424	0.42424	0.42424	0.42424	0.42424	0.40404	0.40404
Doc3	0.38614	0.37624	0.36634	0.29703	0.38614	0.39604	0.39604
Doc4	0.40000	0.40000	0.40000	0.40000	0.40000	0.39048	0.39048
Doc5	0.46154	0.46154	0.46154	0.46154	0.46154	0.48077	0.48077
Doc6	0.63725	0.64706	0.64706	0.64706	0.64706	0.73529	0.73529
Doc7	0.41489	0.41489	0.41489	0.41489	0.41489	0.45745	0.45745
Doc8	0.43299	0.49485	0.49485	0.43299	0.43299	0.46392	0.46392
Doc9	0.39806	0.38835	0.42718	0.51456	0.51456	0.33010	0.33010
Doc10	0.46078	0.46078	0.46078	0.52941	0.52941	0.56863	0.60784

ROUGE-2 recall scores of different techniques and our proposed scoring methods are shown in Table-2. For documents 2,6,7,8,9,10 our proposed method2 performs better as compared to TF, Linear Sum, MEAD, MS Word and LSA. For documents 3,4,5 all methods perform in the similar range. This shows that our proposed method2 is producing more efficient summaries as compared to other approaches.

Table 2. ROUGE-2 recall scores of different techniques and proposed approaches.

Document Number	TF	Linear Sum	MEAD	MS Word	LSA	Proposed Method1	Proposed Method2
Doc1	0.44860	0.44860	0.44860	0.28037	0.31776	0.14019	0.14019
Doc2	0.20388	0.20388	0.20388	0.20388	0.20388	0.23301	0.23301
Doc3	0.19231	0.09615	0.09615	0.08654	0.19231	0.19231	0.19231
Doc4	0.07921	0.07921	0.07921	0.07921	0.07921	0.07921	0.07921
Doc5	0.14679	0.14679	0.14679	0.14679	0.14679	0.14679	0.14679
Doc6	0.36275	0.36275	0.36275	0.36275	0.36275	0.40196	0.40196
Doc7	0.08654	0.08654	0.08654	0.08654	0.08654	0.09615	0.09615
Doc8	0.18447	0.33981	0.33981	0.18447	0.18447	0.28155	0.28155
Doc9	0.10784	0.08824	0.10784	0.13725	0.13725	0.13725	0.13725
Doc10	0.13725	0.13725	0.13725	0.18627	0.18627	0.19608	0.19608

ROUGE-L recall results for proposed methods and other methods used for comparison are presented in Table-3. For document number 2,3,5,6,7,8,10 our proposed method2 performs better as compared to other approaches. For document 4 results are in similar range of all methods.

Table 3. ROUGE-L recall scores of different techniques and proposed approaches.

Document Number	TF	Linear Sum	MEAD	MS Word	LSA	Proposed Method1	Proposed Method2
Doc1	0.61111	0.61111	0.61111	0.50000	0.51852	0.38889	0.38889
Doc2	0.41346	0.41346	0.41346	0.41346	0.41346	0.42308	0.42308
Doc3	0.36190	0.30476	0.30476	0.23810	0.36190	0.37143	0.37143
Doc4	0.27451	0.27451	0.27451	0.27451	0.27451	0.27451	0.27451
Doc5	0.41818	0.40000	0.40000	0.41818	0.41818	0.44545	0.44545
Doc6	0.48544	0.48544	0.48544	0.48544	0.48544	0.53398	0.53398
Doc7	0.32381	0.32381	0.32381	0.32381	0.32381	0.36190	0.36190
Doc8	0.35577	0.49038	0.49038	0.35577	0.35577	0.45192	0.45192
Doc9	0.26214	0.28155	0.26214	0.30097	0.30097	0.25243	0.25243
Doc10	0.27184	0.25243	0.25243	0.34951	0.34951	0.36893	0.37864

Table 4. ROUGE-W recall scores of different techniques and proposed approaches.

Document Number	TF	Linear Sum	MEAD	MS Word	LSA	Proposed Method1	Proposed Method2
Doc1	0.61111	0.61111	0.61111	0.50000	0.51852	0.38889	0.38889
Doc2	0.42308	0.42308	0.42308	0.42308	0.42308	0.44231	0.44231
Doc3	0.36190	0.36190	0.30476	0.23810	0.36190	0.37143	0.37143
Doc4	0.27451	0.27451	0.27451	0.27451	0.27451	0.27451	0.27451
Doc5	0.41818	0.41818	0.40000	0.41818	0.41818	0.44545	0.44545
Doc6	0.48544	0.48544	0.48544	0.48544	0.48544	0.53398	0.53398
Doc7	0.32381	0.32381	0.32381	0.32381	0.32381	0.36190	0.36190
Doc8	0.33654	0.33654	0.47115	0.33654	0.33654	0.43269	0.43269
Doc9	0.25243	0.25243	0.26214	0.30097	0.30097	0.27184	0.27184
Doc10	0.26214	0.26214	0.26214	0.34951	0.34951	0.36893	0.37864

Results of ROUGE-W for proposed and existing methods are shown in Table 4. Similar to ROUGE-1, ROUGE-2 and ROUGE-L, for ROUGE-W also our proposed method performs better as compared to other approaches. Our proposed method2 performs better for document number 2,3,5,6,7,8,9,10. Average scores of all 10 documents using ROUGE-1, ROUGE-2, ROUGE-L and ROUGE-W for proposed and existing approaches used for comparison is presented in Table 5. ROUGE-2 average score is higher due to a single document. However, for rest of the others our proposed method2 gives higher results.

Table 5. Average ROUGE recall scores of different techniques and proposed approaches.

ROUGE Method	TF	Linear Sum	MEAD	MS Word	LSA	Proposed Method1	Proposed Method2
ROUGE-1	0.45355	0.45778	0.46067	0.46315	0.47108	0.47365	0.47757
ROUGE-2	0.19496	0.19892	0.20088	0.17541	0.18972	0.19045	0.19045
ROUGE-L	0.37782	0.38375	0.38180	0.36597	0.38021	0.38725	0.38822
ROUGE-W	0.37491	0.37491	0.38181	0.36501	0.37925	0.38919	0.39017

Average scores show robustness of our method as compared to other methods. Term frequency alone cannot give good results. In real futures when applied individually they cannot produce quality summaries. In case of linear sum the impact of individual feature strength is reduced so that, overall quality of summary degrades. The average score for linear sum in case of ROUGE-1 is 0.45778 and for proposed method2 is 0.47757. In similar way other methods used combination of features. For MEAD, MS Word and LSA the ROUGE-1 scores are 0.26067, 0.46315 and 0.47108 respectively. In case of ROUGE-L and ROUGE-W also the results of our approach is higher. The average score comparisons are shown in Figure 3.

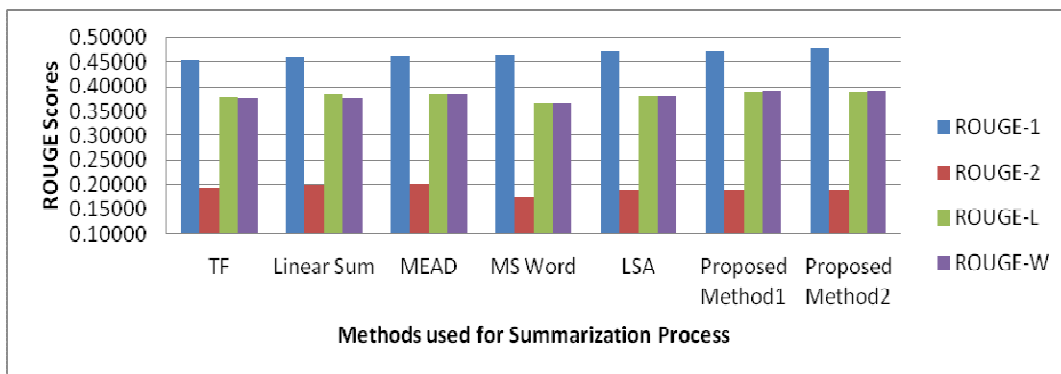


Figure 3. Comparison of average ROUGE scores of proposed and other approaches.

Quality of text summaries changes with selection different set of sentences. As we used scoring methods, our approach used linear combinations in the form of clusters. These clusters in real compute local candidate sentences that can be used for summaries. But combining them furthered linearly degrades the quality of text summaries. Our approach used the inverse values of scores of these clusters that improved the quality of text summaries generated. Experimental results confirm the same. Overall, our proposed method2 outperforms as compared to all other approaches. Performance on individual documents proves uniformity of our results on all evaluation measures without using any kind of semantic processing and inclusion of title feature.

6. CONCLUSIONS AND FUTURE SCOPE

In this paper we investigated the impact of features such as term frequency, gain, sentence location, named entity, LSA, bushy path, sentence centrality, textrank and word co-occurrence. These features are then used to create clusters of local candidate sentences. Their inverse scores are used for final selection of sentences. Our approach alleviates the issue of feature weighing to produce efficient summaries. Furthermore, proposed approach does not require any kind of semantic processing. Title feature is also not included. This makes proposed approach more effective in absence of title of source document. Experimental results on ROUGE-1, ROUGE-2, ROUGE-L and WOUGE-W shows that our proposed scoring performs better as compared to TF, MEAD, MS Word, Linear Sum and modified LSA. In future we will try to analyse the impact of sentence filtering approaches on the quality of text summaries using proposed sentence scoring.

REFERENCES

- [1] K. S. Jones. Automatic summarising: Factors and directions. In *Advances in Automatic Text Summarization*, pages 1-12. MIT Press, 1998.
- [2] E. Hovy and C.-Y. Lin. "Automated text summarization and the summarist system". In *Proceedings of a Workshop on Held at Baltimore, Association for Computational Linguistics.*, Maryland: October 13-15, TIPSTER '98, pages 197-214, Stroudsburg, PA, USA, 1998.
- [3] I. Mani and M. T. Maybury. *Advances in Automatic Text Summarization*. MIT Press, Cambridge, MA, USA, 1999.
- [4] H. P. Luhn. "The automatic creation of literature abstracts". *IBM Journal of Research and Development*, 2(2):159-165, Apr. 1958.
- [5] P. B. Baxendale. "Machine-made index for technical literature: An experiment". *IBM Journal of Research and Development.*, 2(4):354-361, 1958.
- [6] H. P. Edmundson. "New methods in automatic extracting". *J. ACM*, 16(2):264-285, Apr. 1969.
- [7] J. E. Rush, R. Salvador, and A. Zamora. "Automatic abstracting and indexing. ii. production of indicative abstracts by application of contextual inference and syntactic coherence criteria". *Journal of the American Society for Information Science*, 22(4):260-274, 1971.
- [8] C. Lin and E. Hovy." Identifying topics by position". In *Proceedings of the fifth conference on Applied natural language processing (ANLC '97)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 283-290. 1997
- [9] D. Radev, S. Blair-Goldensohn and Z. Zhang. "Experiments in single and multi-document summarization using MEAD". In *First Document Understanding Conference*, New Orleans LA, 2001.
- [10] A. Abuobieda, N. Salim, A. Albaham, A. Osman, and Y. Kumar. "Text summarization features selection method using pseudo genetic-based model". In *International Conference on Information Retrieval Knowledge Management (CAMP)*, pages 193-197, 2012.
- [11] M. A. Fattah and F. Ren. "Ga, mr, _nn, pnn and gmm based models for automatic text summarization". *Computer Speech and Language*, 23(1):126-144, 2009.
- [12] J. J. Pingali, and V. Varma. "Sentence extraction based on single document summarization". In *Workshop on Document Summarization*, 2005.
- [13] M. Mendoza, S. Bonilla, C. Noguera, C. Cobos, and E. Leon. "Extractive single-document summarization based on genetic operators and guided local search". *Expert Syst. Appl.*, 41(9):4158-4169, 2014.
- [14] C. Nobata, S. Sekine, M. Murata, K. Uchimoto, M. Utiyama, and H. Isahara. "Sentence extraction system assembling multiple evidence". In *Proceedings of the Second NTCIR Workshop Meeting*, pages 5-213, 2001.
- [15] J. J. Pollock and A. Zamora. "Automatic Abstracting Research at Chemical Abstracts Service". *Chemical Information and Computer Sciences*, 15(4):226-232, Nov. 1975.
- [16] P. R. Shardanand and U. Kulkarni. "Implementation and evaluation of evolutionary connectionist approaches to automated text summarization". In *Journal of Computer Science*, pages 1366-1376, Feb 2010.
- [17] R. Ferreira, L. D. Cabral, R. D. Lins, G. P. Silva, F. Freitas, G. D.C. Cavalcanti, R. Lima, S. J. Simske and L. Favaro, "Assessing sentence scoring techniques for extractive text summarization",

- Expert Systems with Applications, Volume 40, Issue 14, 15, Pages 5755-5764, ISSN 0957-4174, 2012.
- [18] R. Ferreira., F. Freitas ,L. D. DeSouza, L. Cabral, R. D. Lins,R. Lim, G. Franca.,S. J. Simske and L. Favaro. "A Context Based Text Summarization System". 11th IAPR International Workshop on Document Analysis Systems (DAS), vol., no., pp.66,70, 7-10,2014.
- [19] Y. K. Meena and D. Gopalani, "Analysis of Sentence Scoring Methods for Extractive Automatic Text Summarization", International Conference on Information and Communication Technology for Competitive Strategies (ICTCS-2014), India , ACM 978-1-4503-3216-3/14/11, 2014.
- [20] C. Lin. "ROUGE: a Package for Automatic Evaluation of Summaries". In Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004), Barcelona, Spain, 2004.
- [21] T. Mori. "Information gain ratio as term weight: the case of summarization of ir results". In Proceedings of the 19th International Conference on Computational Linguistics, 688–694. Association for Computational Linguistics Publisher, 2002
- [22] Y. Wang and J. Maches, "A Comprehensive Method for Text Summarization Based on Latent Semantic Analysis," NLPCC, Springer-Verlag Berlin Heidelberg, pp. 394–401, 2013.

Y. K. Meena is Assistant Professor in Department of Computer Science & Engineering in Malaviya National Institute of Technology (MNIT) Jaipur. He was graduated in Information Technology from MNIT, Jaipur in the year 2005. He obtained Master's degree in Computer Applications from IIT Delhi. His research interests include issues related to Data Mining, Natural Language Processing, Pattern Recognition and Knowledge Management. He has published 15 research papers in various International Conferences. He has 08 years of Teaching and Industrial experience. Currently he is pursuing his Ph.D. from MNIT Jaipur.



Dinesh Gopalani received his PhD in Computer Science & Engineering from Malaviya National Institute of Technology, Jaipur, India. He is Assistant Professor at the Department of Computer Science & Engineering Malaviya National Institute of Technology, Jaipur, India from past 15 years. His research interests include issues related to Aspect Oriented Programming, Compiler Design, Natural Language Processing and Knowledge Management. He has published research papers at national and international journals, conference proceedings as well as chapters of books.

