IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-4, Issue-9 E-ISSN: 2347-2693

Strategy for Hindi Text Summarization using Content Based Indexing Approach

Swati Sargule[1*], Ramesh M Kagalkar[2]

[1]Department of Computer Engineering,
[2] Dr. D. Y. Patil School of Engineering and Technology, Savitribai Phule Pune University, Pune, India

Available online at: www.ijcseonline.org

Received: 12/Aug/2016 Revised: 22/Aug/2016 Accepted: 15/Sept/2016 Published: 30/Sep/2016

***Abstract*—** The Document summarization provides summary of document in a very short time. Existing systems for document summarization have work carried on English text summarization. Such systems do not consider the context of the word to produce summary. Previously implemented document summarization models generally use the similarity among sentences in the original document to extract the most relevant sentences. The documents along with the sentences are generally indexed using standard term indexing computation methods, which do not take into account the context related to document. System takes Hindi document as input. That document undergoes through the algorithm and final output is produced as summary of input Hindi document by considering the context of the word. The Bernoulli Model of Randomness technique is used to check the probability of co-occurrences of two terms in large corpus. The methodology used contains lexical association, sentences indexing, word indexing.

***Keywords* —** Document Summarization, Lexical Association, Context Indexing

I. INTRODUCTION

Hindi Document Summarization (DS) is an Information Retrieval (IR) process in which summary of document is extracted to provide brief idea of that document. Existing models on summarization generally use the similarity among sentences in the original document to pick the maximum relevant sentences. The documents along with the sentences are generally indexed using standard term indexing computation techniques, which do not take into account the context related to document. Thus, the similarity values of sentence are independent of the context.

Here a context sensitive document indexing model is considered which is based on the Bernoulli model of randomness for Hindi text document. The Bernoulli model has been used to validate the probability of the co-occurrences of two terms in a large set of documents. Data customers are drowning in natural language text. Whereas the web has enhanced access to text collections on a variety of topics, customers currently face a substantial quantity of redundancy within the texts they encounter online. Document summarization is a data retrieval task, which aims at extracting a concise version of the original document. Single-document summarization and multi-document summarization are terribly closely related tasks and that they are widely investigated independently. Single-document summarization aims to supply a concise and fluent summary for one document, and multi-document summarization aims to supply a concise and fluent summary for a document set consisting of multiple related documents. The two tasks are very closely related in each task definition and answer method. Moreover, each of them is vital in several data systems and applications. As an example, given a cluster of news articles, a multi--document summary are often wont to facilitate users to know the full cluster, and one summary for every article are often -wont to facilitate users to understand the content of the desired article. Multi-document summarization aims to supply a summary describing the most topic in a document set, with none previous information. Multi-document summary are often used to facilitate users to quickly understand a document cluster. As an example, variety of news services have been developed to group news articles into news topics, and so produce a brief summary for every news topic. Users will simply understand the subject they need interest in by taking a glance at the short summary, while not wanting into every individual article within the subject cluster.

This paper is organized as follows: Section 2 discusses the existing document summarization work. In Section 3, we propose Hindi Document summarization system. In Section 4, methodology used in system is discussed. In Section 5, we analyze and compare the proposed schemes in terms of time required to produce summary. Section 6 concludes this paper.

II. LITERATURE SURVEY

Context Based Word Indexing Model is proposed in [1] where Bernoulli Model of Randomness is used for English

Text Summarization to produce relevant summary from input document.

Single Document summarization by exploiting neighborhood knowledge is proposed in [2]. It is based on Key phrase Extraction. The summary is produced from single document. In [4] examines the mutual influences in the two tasks and puts a novel unified approach to simultaneous single-document and multi-document summarizations. The mutual influences in the two tasks are incorporated into a graph model and the ranking scores of sentence for the two tasks can be obtained in a unified ranking process. In [5] aims to explore document impact on summarization performance. They propose a document-based graph model to incorporate information at document-level and the relationship between sentence and document into the graph-based ranking process. In this study, they directly make use of the coarse grained document-level information. A document can be segmented into a few subtopic passages by using the Text Tiling algorithm and according to their knowledge the subtopic passage is more fine-grained than the original document. In [7], proposed a new principled and versatile framework for multi-document summarization using the minimum dominating set. They show that four well-known summarization tasks including generic, query-focused, update and comparative summarization can be modelled as different variations derived from the proposed framework. Approximation algorithms for performing summarization are also proposed and empirical experiments are conducted to demonstrate the effectiveness of their proposed framework.

In [8] proposed the Cluster-based Conditional Markov Random Walk Model (ClusterCMRW) and the Cluster-based HITS Model (ClusterHITS) to fully give the cluster-level information. The first model incorporates the cluster information in the Conditional Markov Random Walk Model and the second model uses the HITS algorithm by considering the cluster as hubs and the sentences as authorities.

In [9] proposed a new multi-document summarization framework based on sentence-level semantic analysis and symmetric non-negative matrix factorization. They first calculate sentence to sentence similarities using semantic analysis and construct the similarity matrix. Then symmetric matrix factorization, which has been shown to be equivalent to normalized spectral clustering, is used to group sentences into clusters. Finally, the most informative sentences are selected from each group to produce the summary. In [10] investigated the impact that methods for topic representation and structuring can have on the quality of multi-document summaries. In [11] propose a technique that, given a keyword query, on the fly produces new pages, called composed pages, which contain all query keywords. The composed pages are generated by getting and putting together related pieces from hyperlinked Web pages and retaining links to the original Web pages. To number the composed pages, we consider both the hyperlink structure of the original pages and the associations between the keywords within each page. Furthermore, they present and experimentally evaluate heuristic algorithms to efficiently generate the top composed pages. In [12] proposed an opinion mining and summarization method using different approaches and resources, valuating each of them in turn. Their work includes the improvement of the polarity classification component by using machine learning over annotated corpora and other techniques, such as anaphora resolution.

In [13] present the first report of automatic sentiment summarization in the legal domain. This work is based on processing a legal questions with a system containing a semiautomatic Web blog search module and FastSum, a complete automatic extractive multi-document sentiment summarization system. They provide quantitative evaluation results of the summaries using legal expert reviewers. They report baseline evaluation results for query-based sentiment summarization. In [14], author proposed a novel algorithm for opinion summarization that takes account of content and coherence, simultaneously. They consider a summary as a sequence of sentences and directly acquire the optimum sequence from multiple review documents by pulling and ordering the sentences. They achieve with a novel Integer Linear Programming (ILP) formulation. The system in this paper is a powerful mixture of the Maximum Coverage Problem and the Traveling Salesman Problem, and is largely applicable to text generation and summarization.

In [15] author defines a task called topic anatomy, which summarizes and relates the core parts of a topic temporally so that readers can understand the content easily. The proposed model, called TSCAN, derives the major themes of a topic from the eigenvectors of a temporal block association matrix. Then, the important events of the themes and their summaries are extracted by examining the constitution of the eigenvectors. Finally, the extracted events are associated through their temporal closeness and context similarity to form an evolution graph of the topic. Paper in [17] provides a comprehensive theoretical analysis of a parametric query vector, which is assumed to show the information needs of the user. A lexical association function has been derived analytically using the system relevance criteria. The derivation is further justified using empirical evidence from the user relevance criteria. They provides mathematical framework to the query expansion techniques, which have largely been heuristic in the existing literature. By using the generalized retrieval framework, the proposed query representation model is applicable to the vectors space model (VSM), Okapi best matching 25 (Okapi BM25), and Language Model (LM).

In [18] the multi-variate Bernoulli model is introduced with respect to its successor and examine its role in future retrieval systems. In the context of Bayesian Learning both modeling approaches are described, contrasted and compared theoretically and computationally. They show that the query likelihood following a multi-variate Bernoulli distribution introduces interesting retrieval features which may be useful for specific retrieval tasks such as sentence retrieval. Then, they addressed the efficiency aspect and show that algorithms can be designed to perform retrieval efficiently for multi-variate Bernoulli models, before making an empirical comparison to study the behavioral aspects of the models. A serial order of comparisons is then conducted on a number of test collections and retrieval tasks to determine the empirical and practical differences between the different models. In [19] describe a sentence position based summarizer that is made based on a sentence position policy, created from the evaluation testbed of recent summarization tasks at Document Understanding Conferences (DUC). In [20] consider another kind of position information, i.e. the word position information, which is based on the ordinal positions of word appearances instead of sentence positions. An extractive summarization model is proposed to give an evaluation framework for the position information. The resulting systems are evaluated on various data sets to demonstrate the effectiveness of the position information in different summarization tasks. In [21-38] propose an alternative way to estimate a translation model based on normalized mutual information between words, which is less computationally expensive and has better coverage of query words than the synthetic query method of estimation. They also propose to regularize estimated translation probabilities to ensure sufficient probability mass for self-translation.

III. PROPOSED SYSTEM OVERVIEW

The proposed system takes Hindi documents as input and the terms encountered in it will either be topical or non-topical. Whereas it's tough to make a decision regarding the interestingness of a term solely on the idea of one document, the patterns of term co-occurrence over a bigger information set are often useful. Lexical association measures use the term co-occurrence information extracted from an outsized domain. Once the lexical association is calculated, construct the document graph, with the terms showing within the document because the vertices and therefore the lexical association between these terms form the edges of the graph. Bernoulli model of randomness has been used to derive a novel lexical association measure. Content based word indexing is used to calculate sensitive indexing weight of each term. Similarly context based sentence indexing is computed for sentence sensitivity indexing.

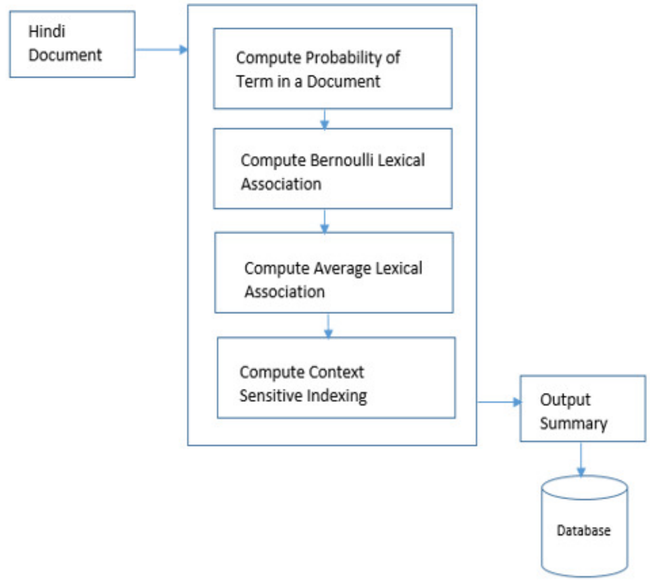

Figure 1. System Architecture

System takes Hindi documents as input and the terms encountered in it can either be topical or non-topical. While it is difficult to decide about the topicality of a term only on the basis of a single document, the patterns of term co-occurrence over a larger data set can be helpful. Lexical association measures use the term co-occurrence knowledge extracted from a large corpus. Once the lexical association is calculated, construct the document graph, with the terms present in the document as the vertices and the edges of the graph as the lexical association between these terms. Bernoulli model of randomness has been used to derive a novel lexical association measure. Content based word indexing is used to calculate sensitive indexing weight of each term. Similarly context based sentence indexing is computed for sentence sensitivity indexing.

IV. METHODOLOGY

Consider a set of S documents. Let these documents have s unique words, which will be used to index these documents, thus called "index terms."

Let $T = \{t_1, t_2, \dots, t_n\}$ be the set of such index terms. Let the set of S documents be $D = \{D_1, D_2, \dots, D_S\}$, Let f_{ij} be the frequency with which term t_j occurs in document D_i and S_j be the number of documents in which the term t_j occurs at least once. S_j is also called the document frequency of term t_j.

Denote the probability of term t_i appearing in the corpus by p_i. Let S_{ij} denote the number of documents in which terms t_i and t_j co-occur.

A. Bernoulli Model of Randomness

Let's assume that the terms to be distributed as per the Bernoulli distribution, the probability pi of the term t_i appearing in a document is given by-

$$p_i = \frac{S_i}{S} \quad (1)$$

Term t_i occurs in S_{ij} documents out of these S_j documents and does not occur in $S_j - S_{ij}$ documents. Therefore, the probability of S_{ij} co-occurrences in S_j documents is given by the Bernoulli distribution-

$$Prob(S_{ij}) = B(S, S_j, S_{ij}) = \frac{s_j}{s_{ij}} p_i^{S_{ij}} q_i^{S_j - S_{ij}} \quad (2)$$

Where $q_j = 1 - p_i$

As per classical semantic information theory, the quantity of information associated is equivalent to the reciprocal of this probability, expressed in bits. Therefore, the information content in the S_{ij} co-occurrences of term t_i in S_j documents can be expressed as-

$$Inf(S_{ij}) = -log_2\left(Prob(S_{ij})\right)$$
$$= -log_2\left(\frac{s_j}{s_{ij}}\left(p_i^{S_{ij}} q_i^{S_j - S_{ij}}\right)\right) \quad (3)$$

Above equation needs the calculation of factorials and thus, Stirling's approximation can be used to approximate the factorials comprised in the calculation. Rendering to Stirling's approximation-

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \quad (4)$$

Proceeding beside the similar lines of the derivation, originate the data content in the S_{ij} co-occurrences as-

$$Inf(S_{ij}) = 0.5log_2\left(2\pi S_{ij}\left(1 - \frac{S_{ij}}{S_j}\right)\right) + S_{ij}log_2\frac{p_{co}}{p_i}$$
$$+(S_j - S_{ij})log_2\frac{1-p_{co}}{1-p_i} \quad (5)$$

Where $p_i = \frac{S_i}{S}$ and $p_{co} = \frac{S_{ij}}{S_j}$

Use this information as the "Bernoulli lexical association measure" to give a context-sensitive indexing weight to the document terms.

If a document has additional target summary, an average of lexical association was taken over the average lexical association for every summary. For instance, if a document or summary has W words (exclusive of the stop words), the average lexical association for the documents or summary was calculated as-

$$avLex = \frac{\sum_{i=1}^{W}\sum_{j\neq i,j=1}^{W} A_{ij}}{W(W-1)} \quad (6)$$

Where, A_{ij} resembles to the lexical association among i^{th} and j^{th} word in the document or summary.

B. Context Based Word Indexing

Next job is to compute the context sensitive indexing weight of every term in a document. A graph-based iterative algorithm is used to search the context sensitive indexing weight of every term. Specified a document D_i, a document graph G is constructed. Let $G = (V, E)$ be an undirected graph to redirect the associations among the terms in the document D_i. $V = \{v_j | 1 \leq j \leq |V|\}$Denotes the set of vertices, where every vertex is a term looking in the document. E is a matrix of dimensions $|V| \times |V|$. Every edge $e_{jk} \in E$ relates to the lexical association value among the terms equivalent to the vertices v_j and v_k. The lexical association among the similar term is set to 0. E is normalized $\tilde{E} = (\tilde{E}_{j,k})_{|V|\times|V|}$to make the sum of each row equal to 1 by means of use the PageRank-based algorithm. $\tilde{E}$is defined as:

$$\tilde{E}_{j,k} = \begin{cases} \frac{e_{jk}}{\sum_{k=1}^{|V|} e_{jk}}, & j \neq k \\ 0, & otherwise \end{cases} \quad (7)$$

Context-sensitive indexing weight of every one word v_j in a document D_i, signified by $indexWt(v_j)$ and a buffer that stores the indexing weights of the previous iteration is signified by $memoWt(v_j)$.

$$ContextBasedWordIndexing(E, \mu, \varepsilon)$$
$$initialize \quad indexWt[v_j] \leftarrow 1 \forall j, \quad error \quad E \leftarrow 1$$
$$while \quad E \leftarrow \varepsilon$$
$$do\begin{cases} E \leftarrow 0 \\ for \quad j \leftarrow 0 \quad to \quad |S| \\ do\begin{cases} memoWt[v_j] \leftarrow indexWt[v_j] \\ indexWt[v_j] \leftarrow \mu.\sum_{\forall k \neq j} indexWt[v_j].\tilde{E}_{kj} \\ \qquad + \frac{1-\mu}{|V|} \\ E \leftarrow E + (indexWt[v_j] - memoWt[v_j])^2 \end{cases} \\ E \leftarrow \sqrt{E} \end{cases}$$
$$return \quad indexWt[v_j]$$

The above defined model gives a context-sensitive indexing weight to every document term. The afterward stage is to use these indexing weights to compute the correspondence among any two sentences. Assumed a sentence st_i in the document D_i , the sentence vector is built using the $indexWt()$. The sentence vector $\overline{st_j}$ is computed such that if a term v_k seems in st_j , it is specified a weight $indexWt(v_k)$; else, it is assumed a weight 0. The similarity among two different sentences st_j and st_i is calculated by means of the dot product, that is $sim(st_j, st_i) = \overline{st_j} \,.\, \overline{st_i}$

V. EXPERIMENT AND RESULTS

The following graph in Figure. 2 shows graph for the time taken by the proposed system for generating Hindi text summary for different sentences count. There are multiple documents varying in in the sentences count and sentence length. The readings of time are shown table 1. System takes nearly 49 seconds to produce summary of 22 sentences. Similarly, system takes nearly 59 seconds to produce summary of 44 sentences. These time comparison is displayed in the below bar graph. X-axis shows number of sentences and Y-axis shows time in seconds.

Number of Sentences	22 Sentences	44 Sentences	46 Sentences	54 Sentences
Time taken to produce summary	49 Seconds	59 Seconds	60 Seconds	62 Seconds

a. Time comparison to produce summary with respect to number of Sentences

b.

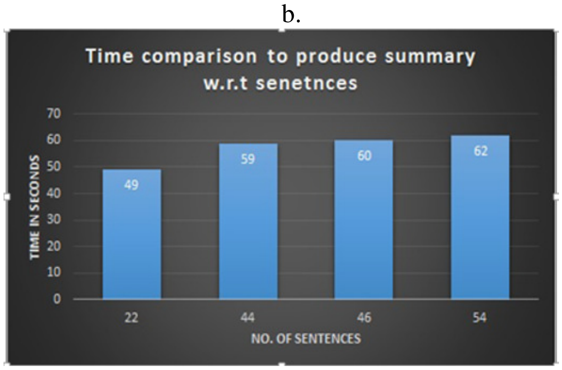

Figure 2. Time comparison to produce summary w.r.t sentences

The following graph in Figure. 3 shows graph for the time taken by the proposed system for generating Hindi text summary for different words count. There are multiple documents varying in in the words count. The readings of time w.r.t words are shown table 2. System takes nearly 45 seconds to produce summary of 161 words. Similarly, system takes nearly 55 seconds to produce summary of 357 words. These time comparison is displayed in the below bar graph. X-axis shows number of Words and Y-axis shows time in seconds.

Number of Words	161 Words	357 Words	369 Words	408 Words
Time taken to produce summary	45 Seconds	55 Seconds	57 Seconds	65 Seconds

b. Time comparison to produce summary with respect to number of Sentences

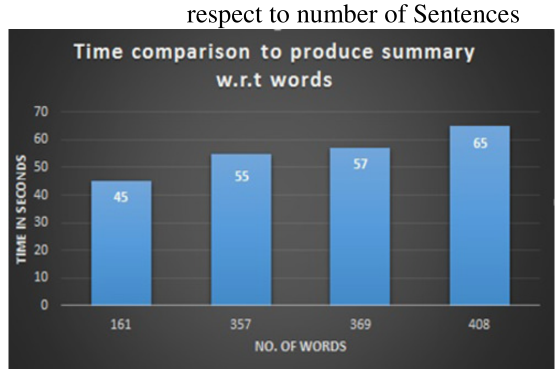

Figure 3. Time comparison to produce summary w.r.t words

VI. CONCLUSION

In this paper, system has introduced an approach for generating summary from Hindi text document by using Bernoulli Model of Randomness. This model has been used to develop a graph-based ranking algorithm for computing how informative each of the document terms is. It is been used to derive a novel lexical association measure which is useful to generate indexing values for each word in the input document. These indexing values has been used to select words for output summary.

System considered only single Hindi text document. In future, we will try to build system that generate summary from multiple Hindi documents. Also we will upgrade system which will support multiple natural languages text like Kannada, Gujrati and Tamil etc.

ACKNOWLEDGMENT

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. I am thankful to the Ramesh M. Kagalkar for his valuable guidance and constant guidelines also thank full the computer department staff of DYPSOET, Lohegoan, Pune and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

REFERENCES

[1] P. Goyal, L. Behera, and T. M. McGinnity, "A Context-Based Word Indexing Model for Document Summarization", IEEE Trans. on Knowledge and Data Engineering, vol. 25, no. 8, August 2013.

[2] Q. Cao and S. Fujita, "Cost-effective replication schemes for query load balancing in DHT-based peer-to-peer P. Goyal, L. Behera, and T. M. McGinnity, "A Context-Based Word Indexing Model for Document Summarization", IEEE Trans. on Knowledge and Data Engineering, vol. 25, no. 8, August 2013.

[3] X. Wan and J. Xiao, "Exploiting Neighborhood Knowledge for Single Document Summarization and Keyphrase Extraction," ACM Trans. Information Systems, vol. 28, pp. 8:1-8:34, http://doi.acm.org/10.1145/1740592.1740596, June 2010.

[4] L.L. Bando, F. Scholer, and A. Turpin, "Constructing Query-Biased Summaries: A Comparison of Human and System Generated Snippets," Proc. Third Symp. Information Interaction in Context, pp. 195-204, http://doi.acm.org/10.1145/1840784.1840813, 2010.

[5] X. Wan, "Towards a Unified Approach to Simultaneous Single- Document and Multi-Document Summarizations," Proc. 23rd Int'l Conf. Computational Linguistics, pp. 1137-1145,http://portal.acm.org/citation.cfm?id=1873781.1873909, 2010.

[6] X. Wan, "An Exploration of Document Impact on Graph-Based Multi-Document Summarization," Proc. Conf. Empirical Methods in Natural Language Processing, pp. 755-762, http://portal.acm.org/citation.cfm?id=1613715.1613811, 2008.

[7] Q.L. Israel, H. Han, and I.-Y. Song, "Focused Multi-Document Summarization: Human Summarization Activity vs. Automated Systems Techniques," J. Computing Sciences in Colleges, vol. 25, pp. 10-20, http://portal.acm.org/citation.cfm?id=1747137.1747140, May 2010.

[8] C. Shen and T. Li, "Multi-Document Summarization via the Minimum Dominating Set," Proc. 23rd Int'l Conf. Computational Linguistics, pp. 984-992, http://portal.acm.org/citation.cfm?id=1873781.1873892, 2010.

[9] X. Wan and J. Yang, "Multi-Document Summarization Using Cluster-Based Link Analysis," Proc. 31st Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 299-306,http://doi.acm.org/10.1145/1390334.1390386, 2008.

[10] D. Wang, T. Li, S. Zhu, and C. Ding, "Multi-Document Summarization via Sentence-Level Semantic Analysis and Symmetric Matrix Factorization," Proc. 31st Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 307-314,http://doi.acm.org/10.1145/1390334.1390387, 2008.

[11] S. Harabagiu and F. Lacatusu, "Using Topic Themes for Multi-Document Summarization," ACM Trans. Information Systems, vol. 28, pp. 13:1-13:47, http://doi.acm.org/10.1145/1777432.1777436, July 2010.

[12] R. Varadarajan, V. Hristidis, and T. Li, "Beyond Single-Page Web Search Results," IEEE Trans. Knowledge and Data Eng., vol. 20,no. 3, pp. 411-424, Mar. 2008.

[13] E. Lloret, A. Balahur, M. Palomar, and A. Montoyo, "Towards Building a Competitive Opinion Summarization System: Challenges and Keys," Proc. Human Language Technologies: The 2009 Ann. Conference of the North Am. Ch. Assoc. for Computational Linguistics, Companion Vol. : Student Research Workshop and Doctoral Consortium, pp. 72-77,http://portal.acm.org/citation.cfm?id=1620932.1620945, 2009.

[14] J.G. Conrad, J.L. Leidner, F. Schilder, and R. Kondadadi, "Query-Based Opinion Summarization for Legal Blog Entries," Proc. 12th Int'l Conf. Artificial Intelligence and Law, pp. 167-176, http://doi.acm.org/10.1145/1568234.1568253, 2009.

[15] H. Nishikawa, T. Hasegawa, Y. Matsuo, and G. Kikui, "Opinion Summarization with Integer Linear Programming Formulation for Sentence Extraction and Ordering," Proc. 23rd Int'l Conf. Computational Linguistics: Posters, pp. 910-918, http://portal.acm.org/citation.cfm?id=1944566.1944671, 2010.

[16] C.C. Chen and M.C. Chen, "TSCAN: A Content Anatomy Approach to Temporal Topic Summarization," IEEE Trans.Knowledge and Data Eng., vol. 24, no. 1, pp. 170-183, Jan. 2012.

[17] B. Andreopoulos, D. Alexopoulou, and M. Schroeder, "Word Sense Disambiguation in Biomedical Ontologies with Term Co-Occurrence Analysis and Document Clustering," Int'l J. Data Mining and Bioinformatics, vol. 2, pp. 193-215, http://portal.acm.org/citation.cfm?id=1413934.1413935, Sept. 2008.

[18] P. Goyal, L. Behera, and T. McGinnity, "Query Representation Through Lexical Assoc. for Information Retrieval," IEEE Trans.Knowledge and Data Eng., vol. 24, no. 12, pp. 2260-2273, Dec. 2011.

[19] D.E. Losada and L. Azzopardi, "Assessing Multivariate Bernoulli Models for Information Retrieval," ACM Trans. Information Systems, vol. 26, pp. 17:1-17:46, http://doi.acm.org/10.1145/1361684.1361690, June 2008.

[20] J. Clarke and M. Lapata, "Discourse Constraints for DocumentCompression," Computational Linguistics, vol. 36, pp. 411-441, 2010.

[21] Y. Ouyang, W. Li, Q. Lu, and R. Zhang, "A Study on Position Information in Document Summarization," Proc. 23rd Int'l Conf. Computational Linguistics: Posters, pp. 919-927, http://portal.acm.org/citation.cfm?id=1944566.1944672, 2010.

[22] M. Karimzadehgan and C. Zhai, "Estimation of Statistical Translation Models Based on Mutual Information for Ad Hoc Information Retrieval," Proc. 33rd Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 323-330, http://doi.acm.org/10.1145/1835449.1835505, 2010.

[23] Swati Sargule, Ramesh M Kagalkar "Hindi Language Document Summarization using Context Based Indexing Model", CiiT International Journal of Data Mining Knowledge Engineering, Vol.08 No. 01, Jan Issue 2016.

[24] Ramesh M. Kagalkar, Dr. Nagaraj H.N and Dr. S.V Gumaste, "A Novel Technical Approach for Implementing Static Hand Gesture Recognition", International Journal of Advanced Research in Computer and Communication Engineering ISSN (Online) 2278-1021 ISSN (Print) 2319-5940 Vol. 4, Issue 7, July 2015.

[25] Amit kumar and Ramesh Kagalkar "Advanced Marathi Sign Language Recognition using Computer Vision" , International Journal of Computer Applications (0975 – 8887) Volume 118 – No. 13, May 2015.

[26] Amit kumar and Ramesh Kagalkar "Sign Language Recognition for Deaf User", Internal Journal for Research in

Applied Science and Engineering Technology, Volume 2 Issue XII, December 2014.
[27] Ramesh. M. Kagalkar and Dr. S. V. Gumaste "Automatic Graph Based Clustering for Image Searching and Retrieval from Database", CiiT Software Engineering and Technology, Vol 8, No 2, 2016.
[28] Ramesh. M. Kagalkar and Dr. S. V. Gumaste "Review Paper: Detail Study for Sign Language Recognition Techniques", CiiT Digital Image Processing, Vol 8, No 3, 2016.
[29] Ramesh M Kagalkar, Kajal Chavan, Asmita Jadhav, Ravina Patil and Asmita Rawool," Self-Educating Tool Kit for Kids" CiiT Software Engineering and Technology, Vol 8, No 1, 2016.
[30] Vasundhara Kadam and Ramesh Kagalkar," Review on Textual Description of Image Contents", International Journal of Computer Applications (0975 – 8887) National Conference on Advances in Computing (NCAC 2015)
[31] Mrunmayee Patil and Ramesh Kagalkar "An Automatic Approach for Translating Simple Images into Text Descriptions and Speech for Visually Impaired People", International Journal of Computer Applications (0975 – 8887) Volume 118 – No. 3, May 2015.
[32] Kaveri Kamble and Ramesh Kagalkar, "Audio Visual Speech Synthesis and Speech Recognition for Hindi Language", International Journal of Computer Science and Information Technologies (IJCSIT) ISSN (Online): 0975-9646, Vol. 6 Issue 2, April 2015.
[33] Shivaji Chaudhari and Ramesh Kagalkar "A Review of Automatic Speaker recognition and Identifying Speaker Emotion Using Voice Signal" International Journal of Science and Research (IJSR), Volume 3, Issue 11 November 2014.
[34] Ajay R. Kadam and Ramesh Kagalkar, "Audio Scenarios Detection Technique", International Journal of Computer Applications (IJCA), Volume 120 June 2015, Edition ISBN: 973-93-80887-55-4.
[35] Ramesh.M.Kagalkar and P.N.Girija" Neural Network Based Document Image Analysis for Text, Image Localization Using Wavelet Decomposition and Mathematical Morphology" International Journal on Computer Science and Information Technology (IJCEIT) Volume 16, No 21, ISSN 0974-2034, Jan-Feb 2010.
[36] Ramesh.M.Kagalkar, Mrityunjaya .V. Latte and Basavaraj.M.Kagalkar "An Improvement In Stopping Force Level Set Based Image Segmentation" International Journal on Computer Science and Information Technology(IJCEIT) ISSN 0974-2034, Volume 24, Issue No 01, June – August 2010.
[37] Vrushali K Gaikwad and Ramesh Kagalkar "Security and Verification of Data in Multi-Cloud Storage with Provable Data Possession", International Journal of Computer Applications (0975 – 8887) Volume 117 – No. 5, May 2015.
[38] Swati Sargule, Ramesh M Kagalkar "Methodology of Context Centered Term Indexing Style Intended For Hindi Language Document Summarization", CiiT International Journal of Software Engineering and Technology, Vol.08 No. 05, June Issue 2016.

AUTHORS PROFILE

Swati Sargule is M.E 2nd year student of Computer Engg. Department, Dr. D. Y. Patil School of Engg. And Technology,Lohegaon, Pune.

Mr Ramesh. M. Kagalkar was born on Jun 1st, 1979 in Karnataka, India and presently working as an Assistant. Professor, Department of Computer Engineering, Dr. D. Y. Patil School Of Engineering and Technology, Charoli, B.K.Via –Lohegaon, Pune, Maharashtra, India. He has 13.5 years of teaching experience at various institutions. He is a Research Scholar in Visveswaraiah Technological University, Belgaum, He had obtained M.Tech (CSE) Degree in 2006 from VTU Belgaum and He received BE (CSE) Degree in 2001 from Gulbarga University, Gulbarga. He is the author of text book Advance Computer Architecture which covers the syllabus of final year computer science and engineering, Visveswaraiah Technological University, Belgaum. One of his research article "A Novel Approach for Privacy Preserving" has been consider as text in LAP LAMBERT Academic Publishing Germany (Available in online). He is waiting for submission of two research articles for patent right. He has published more than 36 research papers in International Journals and presented few of there in international conferences. His main research interest includes Image processing, Gesture recognition, speech processing, voice to sign language and CBIR. Under his guidance four ME students awarded degree in SPPU, Pune, five students at the edge of completion their ME final dissertation reports and two students started new research work and they have published their research papers on International Journals and International conference.

