

A Systematic and Exhaustive Review of Automatic Abstractive Text Summarization for Hindi Language

Amita Garg
Parul Institute of Management and Research
Parul University
Vadodara, India
amita.garg@paruluniversity.ac.in

Dr. Jatinderkumar R. Saini
Narmada College of Computer Application,
Bharuch, India
saini_expert@yahoo.com

Abstract: Text summarization is the process of extracting salient information from the source text and to present that information to the user in the form of summary. It is very difficult for human beings to manually summarize large documents of text. Automatic abstractive summarization provides the required solution but it is a challenging task because it requires deeper analysis of text. Abstractive summarization methods are classified into two categories i.e. structure based approach and semantics-based approach. In this paper, an exhaustive and systematic review of semantics based abstractive text summarization methods for Hindi language has been presented. The main idea is to explore and understand the development done so far both at the international front as well as at Indian level. Besides the main idea, the strengths and weaknesses of each method have also been highlighted. This review has helped in clearly inferring that although abstractive summarization methods produce highly coherent, cohesive, information with less redundant summary, lot of research gap is still there. Finally, it is concluded from the literature studies that there is lots of research scope for abstractive based text summarization for Hindi language as well as for other regional languages of India

Keywords: Text Summarization, Abstractive Summary, Extractive Summary, Semantic analysis.

I. INTRODUCTION

Now a day's summaries are around us like anything, without the headlines news papers seem incomplete! The Movies and books are often accompanied by either reviews or blurbs. Most research articles often begin with abstracts similarly search results also provide the snippets for every searched result. Thus, short summaries offer a great potential to the lawyers, journalists, students as well as to the casual browsers of the Internet.

While giving the short introduction about the topic, summaries also provide the users the evaluation about the relevance of a document without a need to read the whole text. A user can immediately recognize the relevance of the summary as it is not cluttered completely. The displayed summaries in the search result can be considered as an informative tool for the user e.g., summaries could be helpful in digital libraries. It could be useful for the readers of the journals or books to find the relevant information easily. For the general users of the internet, web summaries could be helpful for removing the irrelevant web pages.

With the ease of access and availability of the in-Internet networks, PC software and hardware tools, it has become easier to create the contents that can be shared, stored, written and published. When such a vast material is available at hands, it becomes harder to search as well as find the relevant data. Though search engines can be used for this task, they are unable to semantically analyze or understand the structure of the text. Even if a query is raised the matches for that will depend mostly on frequency of the word. Thus, one may end up trying to scan the huge number of documents. In this regard, the summaries aid users to evaluate the relevant document without a need to read the full text.

With this rapid and huge growth of the online information research and development in Automatic Text Summarization (ATS) is grabbing lot of importance. ATS gives the condensed form of the source information to the user in a need and task

sensitive manner. During this, the summary generated should not be cluttered with details and also should be relevant and recognizable.

With these criteria, summarization is considered as one of the hard problems of NLP because, to carry out it accurately, it is required to understand the point of a text thoroughly. [Hovy and Lin, 1998] defines summary as 'A summary is a text that is produced from one or more texts, that contains a significant portion of the information in the original text(s) and that is no longer than half of the original text(s).' This definition characterizes three important aspects of automatic summary generation

- A Summary can be generated from a single document or a collection of documents.
- In the generated summary, the salient information from the original text should be kept intact.
- Summaries are preferred to be shorter than the original source.

II. APPROACHES TO AUTOMATIC TEXT SUMMARIZATION

Text summarization approaches can be typically split into two groups

- Extractive Text Summarization Extractive summarization takes out the important sentences or phrases from the original documents and groups them to produce a text summary without any modification in the original text. Normally the sentences are in sequence as in the original text document.
- Abstractive Text Summarization Abstractive summarization is performed by understanding the original text with the help of linguistic method to understand and examine the text. The objective of abstractive summarization is to produce a generalized summary, which conveys information in a precise way that gen-

erally requires advanced language generation and compression techniques. Figure 1 [18] shows the different techniques of Abstractive Text summarization method.

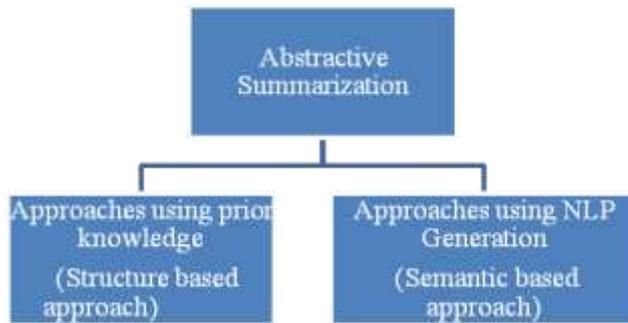


Figure 1. Overview of Abstractive Summarization

Structure Based Approach

In structure-based approach, important sentences from source text gets populated in a predefined structure to obtain the required abstract summary without losing its meaning. The predefined structures used in this approach are templates, tree-based structure, ontology-based structure, lead and body phrase structure and rule-based structure. Figure 2 [18] shows different method of structure-based approach

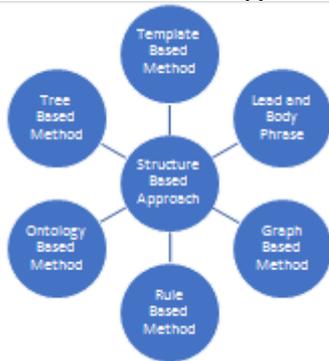


Figure 2. Overview of Structure Based Abstractive Text Summarization

Semantic Based Approach

In semantic based technique, linguistics illustration of document(s) is employed to feed into natural language generation (NLG) system. This technique specializes in identifying noun phrases and verb phrases by processing linguistic data. Figure 3 [18] shows different method of structure-based approach

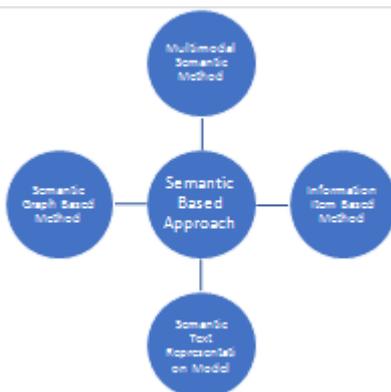


Figure 3. Overview of Semantic Based Abstractive Text Summarization

This article presents a deeper review of different techniques that generate Abstract summaries specially using semantic based method. Further the article is organized as follows. Section II discusses the literature review of various types of Abstractive summarization techniques. Section III presents conclusion and future scope of research.

III. LITERATURE REVIEW OF ABSTRACTIVE TEXT SUMMARIZATION

3.1 International Languages

- a) Piere-Etienne & Guy Lapalme, 2011 [24] the abstractive summarization process initially extracts the abstractive tokens named Information Items (INIT) [24] from the sentences in the document. INIT's are a smallest element of coherent information especially Subject, verb and Object triples [8] appended with date and location elements mentioned in the sentences. These are obtained employing MINIPAR [7] parser and General Architecture for Text Engineering (GATE) [5] information extraction engine. Frequency based model is used for selecting most important INIT's. Simple Natural Language Generator (NLG) [2] considers INIT's as input and accomplish sentences. A constraint is applied in sentence generation; only one INIT is used for generation of sentence. Document frequency score is the bench march for selecting the sentences for final summary. The sentences with highest average document frequencies are selected. Redundancy is eliminated by not considering the lemmas that are already the part of summary. A final step in generating the summary is all the sentences with same known dates are grouped in a single coordinate sentence and "On date" element is included as pre-modifier. The "in location" is appended as post modifier to generated sentences if the INIT is with known location.
- b) (Siddhartha, Prasenjit & Kazunari, 2016) [27] Proposed a procedure for generating generic abstract summary from multi documents. First, distinguish the important document from the given multi document set and each sentence from the important document is deemed as root of the cluster. Second, generating most informative and linguistically structured sentences from each cluster and project them as final summary. LexRank [9], Pair wise Cosine Similarity and Overall Document Collection Similarity techniques are utilized in identifying the important document. The document with highest score among all the three techniques is considered as important. Clusters are formed for similar sentences accounting each sentence from important document as root and finding the similarity with cluster roots among all the sentences from other documents. Generating summary, a word graph [11] is constructed to the sentences for each cluster i.e. one-word graph to one cluster. Nodes in the graph are words along with parts of speech. The node is connected to another node which is adjacent in the sentence with directed edge. K-shortest paths are extracted from the start node to the end node of the sentence. Utilizing Integer Linear Pro-

gramming (ILP) [27] the aggregation of information content value with linguistic quality value a best path is selected from k shortest paths with the objective of most informative and high linguistic quality. Text Rank [26] and Trigram model are the measuring techniques of Information content and Linguistic quality. The constraints of ILP are from each cluster of redundant sentences maximum one sentence is considered in final summary.

- c) (Fei Liu, Jeffrey, Sam, Norman, & Smith, 2015) [10] An Abstractive text summarizer avails graph model to generate generic summaries. The sentences from the documents are organized into Abstractive Meaning Representation (AMR) [17] graphs using JAMR parser [15]. The AMR graphs are semantically represented as direct acyclic graphs with labelled edges. The nodes of the graph are labelled with concepts (entities and events) and edges with relations. The root node of the graph is event. These AMR graphs merge utilizing concept merging technique and graph expansion to construct summary graph. Collapsing of concepts uses AMR concepts which are knowledge base of core semantic relations and role sets adopted from PropBank [14]. Before sub graph selection, graph expansion is done to improve the coverage of information by adding possible edges between nodes within the same AMR graph and across multiple AMR graphs while constructing Source graph. Integer Linear Programming (ILP) [6] taking node level and edge level features as inputs selects sub graph from the summary graph. AMR to text generator is not available. The author stated a heuristic approach is applied to generate a bag of words with no order. This summarizer is termed as graph-to -graph summarizer.
- d) (Riya & Yogesh, 2017) [25] Abstractive approach extracts important content from the document and generates sentences based on the extract as abstractive summary. It employs natural language processing techniques, extraction rules, sentence generation patterns and domain knowledge in summary generation process. This abstractive summary approach is restricted to single domain strictly to news articles on disaster or accident. The Stanford NLP parser [19] splits document into sentences, generates dependency relationships between the tokens in sentence and Named Entity Recognizer (NER) [23] annotates the tokens if tokens are person names, location names, company names, date and time. Extraction rules are created manually based on the dependencies. Using these rules, domain knowledge and dependency relations essential content is extracted. Sentence Generation patterns are employed on extracted content to generate summary. The author states abstractive summarization outperforms INIT [24] based and Template based methods [29]. INIT based methods discards much information in generating short summaries and template-based methods lack of diversity of information.
- e) Moawad& Mostafa, 2012 [21] A graph-based approach uses graph reduction technique to generate abstractive summary to the original document. This

summarizer is organized into three phases Semantic Graph Generation, Graph Reduction, Summary Generation. In Semantic Graph Generation module, deep syntactic analysis of each sentence i.e. dependency relations among words in each sentence, NER, morphological analysis [16], co-reference resolution [28], pronominal resolution [16] are carried on input sentences to pre-process employing Lingsoft [1] and Stanford nlp core parser tools. For each pre-processed sentence a rich semantic sub-graph is generated considering Word Senses Instantiation (using domain ontology the concept of each word either the word is a noun or verb is retrieved), Concepts Validation (based on the senses of each word the sentence concepts are derived by relating and validating), and Semantic Sentences Ranking processes (WordNet usage popularity which calculates concept weights in sentences is utilized to rank sub graphs and keep the semantic consistency in the sentence). Each word in the sentence is represented as triple sequence (stem, morphological and syntactical tags, and dependency relations). The Rich Semantic Graph [22] [3] is generated by merging highest ranked sub-graphs. Graph Reduction phase, implementing heuristic rules, node types, similarities and WordNet [12] [20] semantic relations hyponym, hyponym, entitlement the graph is reduced by merging, deleting and consolidating graph nodes. Summary Generation phase is partitioned into four modules Text planning; Sentence planning, Realization and Evaluation. These modules perform process sequentially resulting final summary. Text planning module takes reduced graph as input and outputs the objects that should include in final summary considering semantic information associated with nodes of the graph. Sentence planning module has four main processes, Lexicalization process, Discourse Structuring Process, Aggregation Process [13], and Referring Expression Process. It considers the output of Text planning module and outputs semi-paragraphs by reducing redundant information. This module achieves smooth flow among sentences. Realization, this module employs Simple NLG [2] [4] tool which outputs paragraphs grammatically correct. Evaluation module calculates the rank of each paragraph accounting most frequently used paragraph word synonyms and coherence.

- f) C. F. Greenbacker [30], proposed a framework for generating an abstractive summary from a semantic model of multimodal document. Multimodal document contains both text and images. The framework has threesteps in first step, a semantic model is constructed using knowledge representation based on objects (concepts) organized by ontology. In second step, informational content (concepts) is rated based on information density metric. The metric determines the relevance of concepts based on completeness of attributes, the number of relationships with other concepts and the number of expressions showing the occurrence of concept in the current document. In third step, the important concepts are expressed as sentences. The expressions observed by the parser are stored in a semantic model for expressing concepts

and relationship. An important advantage of this framework is that it produces abstract summary, whose coverage is excellent because it includes salient textual and graphical content from the entire document. The limitation of this framework is that it is manually evaluated by humans. An automatic evaluation of the framework is desirable.

- g) P.E. Genest and G. Lapalme [31] proposed a framework for abstractive summarization took place in the context of Text Analysis Conference (TAC) 2010 for multi-document summarization of news. The framework consists of following modules Information Item retrieval, sentence generation, sentence selection and summary generation. In Information Item (INIT) retrieval, first syntactic analysis of text is done with parser and the verb's subject and object are extracted. So, an INIT is defined as a dated and located subject-verb-object triple. In sentence generation module, a sentence indirectly generated from INIT using a language generator, the NLG realizeSimpleNLG [22]. Sentence selection module ranks the sentences generated from INIT based on their average Document Frequency (DF) score. Finally, a summary generation step account for the planning stage and include dates and locations for the highly ranked generated sentences. The major strength of this approach is that it produces short, coherent, information rich and less redundant summary. This approach has several limitations. First, many candidate information items are rejected due to the difficulty of creating meaningful and grammatical sentences from them. Secondly, linguistic quality of summaries is very low due to incorrect parses.
- h) Alok Ranjan Pal, Diganta Saha [32] said that the text Summarization is the procedure by which the significant portions of a text are retrieved. Most of the approaches perform the summarization based on some hand tagged rules, such as format of the writing of a sentence, position of a sentence in the text, frequency of few particular words in a sentence etc. But according to different input sources, these predefined constraints greatly affect the result. The proposed approach performs the summarization task by unsupervised learning methodology. The importance of a sentence in an input text is evaluated by the help of Simplified Lesk algorithm. As an online semantic dictionary Word Net is used. First, this approach evaluates the weights of all the sentences of a text separately using the Simplified Lesk algorithm and arranges them in decreasing order according to their weights. Next, according to the given percentage of summarization, a particular number of sentences are selected from that ordered list. The proposed approach gives best results up to 50% summarization of the original text and gives satisfactory result even up to 25% summarization of the original text.
- i) D.Y. Sakhare, Dr. Raj Kumar [33] said that recently, there has been a significant research in automatic text summarization using feature-based techniques in which most of them utilized any one of the soft computing techniques. But, making use of syntactic structure of the sentences for text summarization has not

widely applied due to its difficulty of handling it in summarization process. On the other hand, feature-based technique available in the literature showed efficient results in most of the techniques. So, combining syntactic structure into the feature-based techniques is surely smooth the summarization process in a way that the efficiency can be achieved.

- j) Manisha Prabhakar, Nidhi Chandra [34] presented a paper based on pragmatic analysis. In this paper, text summarization technique is designed for the documents having the fixed format. The proposed system generates the summary of the fixed format documents by analysing all the different parts of the documents. The system consists of five stages. In first stage each sentence is partitioned into the list of tokens and stop words are removed. In second stage, frequency usage is counted for each word. In third stage, assign POS tag for each weighted term and Word sense disambiguation is done. In the fourth stage, pragmatic analysis is performed. After Pragmatic Analysis, summarized sentences will be store in a database.

Indian Languages other than Hindi

- a) Jagadish S Kallimani, Srinivasa K G and Eswara Reddy B [37] proposed a fully abstractive approach for Kannada language with a separate process for the analysis of the text, the content selection, and the generation of the summary. The whole process is divided into 4 steps. Pre-processing performs POS Tagging and stemming on input text document. It also recognises named entities using some defined rules. Next step is Categorization which helps to identify category of text document using statistical methods like TF. After this step attribute extraction is done using IE rules and lastly using an appropriate template, summary is generated. The methodology has proved to have good precision values apart from maintaining readability. Though instrumental in achieving the desired results, usage of templates can bring in flatness or monotony to the summaries generated by this method.
- b) Rajina Kabeer, Sumam Mary Idicula [38] proposed detailed semantic processing of the document to generate the summary. In the Semantic graph based method; the input document first undergoes some linguistic processing in order to extract semantic triples (subject-predicate-object) from each of the sentence in the document. A semantic graph is generated using these semantic triples. A sub graph representing the summary sentence is created by assigning weights to each edges of the graph based on certain features. Those edges with weight less than average weight are deleted from the graph and reduced sub graph is obtained. Summary is generated from this reduced graph. Performance evaluation of the summary generated by this method is compatible with the summary generated by extractive method.
- c) Banu [35] proposed another technique for summarizing documents of Tamil by using approach of sub graph for selecting lines from source document treated as text summary or another technique for generating a generic summary of document. In this system, syntax of language neutral, which is the system

for representing the natural language lines, has been applied for compressing the text documents. It has used syntactic analysis of the source text which makes an analysis of logical form has been used for every line. Triples of subject object predicate are selected from individual lines to generate a semantic graph of source document and its corresponding summary generated by human experts. To triples of SOP Semantic Normalization is used for re Semantic Normalization is used for reducing the frequency of nodes of semantic graph of source document. Classifier has provided training by using leaning technique based on support vector machine learning, for identifying triples of SOP from semantic graph of document which belongs to actual summary. Then this classifier is used to extract automatic summaries from test documents

- d) Keyan [36] proposed multi-lingual (Tamil and English) multi-document summarization by neural networks. The system involves three steps. In first step, the sentences of the documents are converted into vector form. In the second step weight values are assigned to vector form based on sentence features. Depend on sentence weight value, single document summarization is done. The output of single document summarization is used as an input for multi-document Summarization. Final step is a sentence selection, in which output summary is selected based on the similarity and dissimilarity measures. Sentence similarity and dissimilarity measures are used to compare the sentences. From that, resultant summary is produced. The proposed system can be able to summarize both Tamil and English online newspapers.
- e) Varsha R Embar , Surbhi Deshpande [43] in 2013 proposed an abstractive approach to summaries the kanada language text Here the sAramsha system is involve which analyse the document perform the pos (part of speech) in pos it is the words from text is categorized in part in such a way like noun, verb adjective after pos the streaming operation is performed in which stemmers are removed from word then the NER tool which label sequence of words in text which are name of things such as person and company name, organization locations etc IE rules are used to extract the sentences a and then generate the summary.
- f) Deepali K. Gaikwad and Namrata Mahender C. [44] proposed text summarization using questions words as rule to extract the important aspect of the given source text. The system transforms declarative sentences into its interrogative counterpart. The proposed method is focus to generate question that accepts Marathi text as input and processes the input by applying POS tagging, NER, stemming and rule-based approach then generate the question as per the proposed rules. For further use, generated questions are classified and then rank the sentence according to frequency and noun-based priority. The answer of the higher ranked question and noun-based priority questions are group together to form the summary of the

given input. This is the new approach of text summarization.

Hindi Language

- a) Dawinder Kaur, RajbhupinderKaur [42] proposed a rule-based approach for Hindi text abstractive summarization. Handcrafted rules are developed to generate the summary. A corpus of Hindi language named entities is used along with these rules to extract important lines from a text paragraph. A line not having minimum weight of named entity (here it is 5) is not considered for summary. As a second step for summarization, dead phrase removal is also implemented. In this approach dead phrase (combination of two or more words that can be replaced by single word or can be removed) are to be removed from the text paragraph to obtain the more precise results. Dead phrases along with the replacement words are stored in the corpus and hence removed with the help of this corpus from Hindi text paragraph. The proposed system gives an overall accuracy of 91% with 60-70 % compression. If the minimum value for weights of Named Entity is set to 7-8 then compression could be further increased.
- b) Vipul Dalal & Latesh Malik [40] proposed an approach for automatic summarization of Hindi text documents using bio-inspired computing. The paper mainly focuses on pre-processing, machine learning and summary evaluation phases of summarization process. In pre-processing step, the input document is first parsed to give Parts of Speech (POS) tag, and dependency tag. After pre-processing, number of features (word frequency, sentence position, SOV tag, POS tag, page rank, Number of in-coming links, no of outgoing links) from each category is combined to form feature vector of each SOV triples. These SOV triples are labelled as either 0 or 1. Finally a machine learning Particle Swarn Algorithms is used to train a classifier using these labelled triples. This summarization technique gives a very high precision.
- c) Manjula Subramaniam, Prof. Vipul Dalal [39] proposed an approach that aims to summarize an input single text document by creating a semantic graph called Rich Semantic Graph (RSG) for the original document, reducing the generated semantic graph, and then generating the final abstractive summary from the reduced semantic graph. The approach consists of three phases: The Rich Semantic Graph Creation Phase, The Rich Semantic Sub Graph Reduction Phase, and Summarized Text Generation Phase. In RSG, the verbs and nouns of the input document are represented as graph nodes along with edge corresponding to semantic and topological relations between them.

The Rich Semantic Graph Reduction Phase aims to reduce the generated rich semantic graph of the source document to more reduced graph. A model of heuristic rules is applied to reduce the graph by replacing, deleting, or consolidating the graph nodes using the WordNet relations. Finally, the Summarized Text Generation Phase to generate the abstractive summary from the reduced rich semantic graph. This

phase accepts a semantic representation in the form of RSG and generates the summarized text.

- d) Poonam Kolhe, Kumbhare [41] proposed a new approach to text summarization. In this approach first read the source document and then finds the action words with the help of NLP. Then select the lines L1..... Ln, after line selection finds the summary and check using the proposed algorithm given in the paper. Final output of algorithm is the required summary. This system provides an efficient technique for Text Summarization ensures good coverage and avoids redundancy

IV. IDENTIFICATION OF RESEARCH GAPS AND SCOPE FOR RESEARCH

Here we focused on abstractive summarization, which is a challenging research area because of the complexity of natural language processing. Some of the research issues in abstractive summarization methods that need to be addressed have been identified from the review of literature and are presented as follows:

- Scarcity of work done in the field of abstractive summarization in Indian Language as compared to other international language.
- NLU (Natural Language Understanding) and NLG (Natural Language Generation) in Hindi language is an important issue
- Context Identification is again a big area of research as far as Hindi language is concerned.
- There is no generalized framework that humans can use for abstractive summarization.
- Besides synthesizing important sentences for abstractive summary, sentence ordering in a summary is an important research issue.
- Ongoing research on abstractive summarization must still deal with issues such as scarcity of training data, appropriate integration of syntax even when the input data comes from a noisy genre, and compressions involving lexical substitution and paraphrase.
- Compression involving lexical substitution, paraphrasing and reformulation is difficult
- Evaluating an abstractive summary is a difficult task because there does not exist an ideal summary for a given document or set of documents and therefore is an open research

V. CONCLUSION

We have discussed an overview of the existing text summarization techniques in NLP. We have found that text summarization is still a raw area of research as seen many problems found in their produced summaries. Finally, it is concluded from the literature studies that most of the extractive summarization methods produce summaries easily but produce redundant summaries and many of them use WorldNet which is a lexical database for giving sense of a word but still don't give accurate results as it contains limited information and abstractive summarization methods produces highly coherent, cohe-

sive, information rich and less redundant summary but difficult to produce as they need NLG which itself is a growing field. Since the selection of the right procedure of removing problems occurred in both techniques plays an important role, it is important to experiment and a hybrid technique needs to be developed which uses both techniques to get meaningful summary.

REFERENCES

- [1] (2012, June 15) Retrieved from ENGCG Constraint Grammar Parser of English <http://www2.lingsoft.fi/cgi-bin/engcg>.
- [2] A.G. & E. R. (2009). SimpleNLG a Realization Engine for Practical Applications. ENLG'09 proceedings of the 12th European Workshop on Natural Language Generation (pp. 90-93). Morristown, NJ, USA Association for Computational Linguistics.
- [3] Aref, M., Moawad, I., & I. S. (2010). Rich Semantic Graph Generation System Prototype. The 10th Conference on Language Engineering. Cairo, Egypt.
- [4] Carenini, G., & Chi Kit, J. C. (2008). Extractive vs. NLG-based abstractive summarization of evaluative text the effect of corpus controversialist. Proceedings of the 5th International Natural Language Generation Conference (INLG).
- [5] Cunningham, H., Maynard, D., Bontcheva, K., & Tablan, V. (2002). GATE: A framework and graphical development environment for robust NLP tools and applications. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia, PA, USA.
- [6] D. G. & B. F. (2009). A Scalable global model for summarization. Proceedings of the NAACL workshop on Integer Linear Programming for Natural Language Processing.
- [7] Dekang Lin. (1998). Dependency-based evaluation of minipar. Workshop on the Evaluation of Parsing Systems. Granada.
- [8] Delia Rusu, Lorand Dali, B. F., M. G., & D. M. (2007). Triplet extraction from sentences. Proceedings of the 10th International Multiconference "Information Society-IS2007", (pp. 218-222).
- [9] Erkan, G. & Radev, D. R. (2004). LexRank Graph-based Lexical Centrality as Saliency in Text Summarization. Journal of Artificial Intelligence Research (JAIR), 457-479.
- [10] Fei Liu, J. F., S. T., N. S., & Smith, N. A. (2015). Toward Abstractive Summarization Using Semantic Representations. Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics Human Language Technologies (pp. 1077-1086). School of Computer Science at Research Showcase@CMU.
- [11] Filippova, K. (2010). Multi-Sentence Compression Finding Shortest Paths in Word Graphs. Proceedings of the 23rd International Conference on Computational Linguistics (Cooling 2010), (pp. 322-330).
- [12] G. A., N. R., & V. P. (2003). "The Onto WordNet Project Extension and Axiomatization of Conceptual Relations in WordNet". Proceedings of International Conference on Ontologies, Databases and Applications of Semantics (ODBASE 2003), (pp. 820-838). Catania, Italy.
- [13] H. D., & E. H. (2004). Aggregation in Natural Language Generation. EWNLG-93 proceedings of the 4th European Workshop on Natural Language Generation. Pisa, Italy.

- [14] Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., & Weischedel, R. (2006). OntoNotes The 90% solution. Proceedings of NAAC.
- [15] Jeffery, F., Thomson, S., Carbonell, J., Dyer, C., & Smith, N. A. (2014). A Discriminative Graph-Based Parser for Abstract Meaning Representation. Proceedings of ACL.
- [16] L. S., & Leass, H. (1994). An Algorithm for pronoun anaphora resolution. Computational Linguistics, (pp. 535-561).
- [17] Laura, B., Bonial, C., Cai, S., Georgescu, M., Kira, G., Hermjakob, U. Schneider, N. (2013). Abstract meaning representation for sembanking. Proceedings of Linguistic Annotation Workshop.
- [18] N. Moratanch, S. Chitrakala, "A survey on abstractive text summarization", Circuit Power and Computing Technologies (ICCPCT) International Conference on, 2016.
- [19] Manning, D. C., Surdeanu, Mihai, Bauer, John, McColosky David. (2014). The Stanford CoreNLP Natural Language Processing Toolkit. Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics System Demonstrations, (pp. 55-60).
- [20] Miller. (n.d.). Nouns in WordNet A Lexical Inheritance System. Princeton University.
- [21] Moawad, I. F., & M. A. (2012). Semantic Graph Reduction Approach for Abstractive Text Summarization. (pp. 132-138). IEEE.
- [22] Moawad, I., Aref, M., & Ibrahim, S. (2011). Ontology-based Model for Generating Text Semantic Representation. International Journal of Intelligent Computing and Information Sciences, "IJICIS", 117-128.
- [23] Nenkova, A., & McKeown, K. (2003). References to Named Entities A Corpus Study. HLT-NAACL.
- [24] P.-E. G., & G. R.-D. (2011). Framework for Abstractive Summarization using Text-to-Text Generation. Proceedings of 49th Annual Meeting of the Association for Computational Linguistics (pp. 64-73). Portland, Oregon Association for Computational Linguistics.
- [25] R. J. & Y. K. (2017). An Abstractive Approach for Text Summarization. International Journal of Advanced Computational Engineering and Networking, (pp. 5-10).
- [26] Rada, M., & Paul, T. (2004). Text Rank Bringing Order into Texts. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004), (pp. 404-411).
- [27] S. B., P. M., & K. S. (2016). Multi-Document Abstractive Summarization Using ILP based Multi-Sentence Compression. Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (pp. 1208-1214). IJCAI.
- [28] V. N., & C. C. (2002). Improving machine learning approaches to co-reference resolution. Proceedings of 40th Annual Meeting on Association for Computational Linguistics (pp. 104-111). ACL'02.
- [29] Zhou, L., & Hovy, E. (2004). Template-filtered headline summarization. Proceedings of the ACL Workshop on Text Summarization.
- [30] C. F. Greenbacker, "Towards a framework for abstractive summarization of multimodal documents," ACL HLT 2011, p. 75, 2011.
- [31] P.E. Genest and G. Lapalme, "Framework for abstractive summarization using text to-text generation," in Proceedings of the Workshop on Monolingual Text-To-Text Generation, 2011, pp. 64-73.
- [32] Alok Ranjan Pal, Diganta Saha, "An Approach to Automatic Text Summarization using Word Net", IEEE International Conference on Advanced Communication Control and Computing Technologies, 978-1-4799-2572-8/14
- [33] D.Y. Sakhare, Dr. Raj Kumar, "Syntactic and Sentence Feature Based Hybrid Approach for Text Summarization", I.J. Information Technology and Computer Science, Page no- 38-46, 2014
- [34] Manisha Prabhakar, Nidhi Chandra, "Automatic Text Summarization Based On Pragmatic Analysis", International Journal of Scientific and Research Publications, Volume 2, Issue 5, May 2012.
- [35] M. Banu, C. Karthika, P Sudarmani and T.V. Geetha, "Tamil Document Summarization Using Semantic Graph Method", Proceedings of International Conference on Computational Intelligence and Multimedia Applications, pp. 128-134, 2007.
- [36] M.. K. Keyan and K.G. Srinivasagan, "Multi-Document and Multi-Lingual Summarization using Neural Networks", Proceedings of International Conference on Recent Trends in Computational Methods, Communication and Controls, pp. 11-14, 2012.
- [37] J. S. Kallimani, K.G. Srinivasa and B. R. Eswara, "Information Retrieval by Text Summarization for an Indian Regional Language," In Proceedings of International Conference on Natural Language Processing and Knowledge Engineering
- [38] Rajina Kabeer and Sumam Mary Idicula, "Text Summarization for Malayalam Documents – An Experience", in IEEE International Conference on Data Science and Engineering (ICDSE), 2014
- [39] Subramaniam, M., Dalal, V. 2015. Test Model for Rich Semantic Graph Representation for Hindi Text using Abstractive Method in Volume: 02 Issue: 02 May-2015, e-ISSN: 2395-0056
- [40] Dalal, V., Malik, L., 2017. Semantic Graph Based Automatic Text Summarization for Hindi Documents Using Particle Swarm Optimization, International Conference on Information and Communication Technology for Intelligent Systems, pp 284-289.
- [41] Poonam Kolhe, Prof. Ashish Kumbhare, "Optimizing Accuracy of Document Summarization Using Rule Mining" in International Journal of Computer Science and Mobile Computing (IJCSMC), 2017, pp 207-216
- [42] Dawinder Kaur, Rajbhupinder Kaur, "Automatic Summarization of Text Documents Written in Hindi Language" in International Journal of Computer Science and Mobile Computing (IJCSMC), October 2014 pp 207-216
- [43] Varsha R Embar, Surbhi Deshpande "sArAmsha - a Kannada Abstractive Summarizer" 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)
- [44] Deepali K. Gaikwad and Namrata Mahender C., "Rule based Text Summarization for Marathi Text", Journal of Global Research in Computer Science (JGRCS), 2018

AUTHOR'S BIOGRAPHIES



Amita Garg, MCA, working as Assistant Professor at Parul University, Vadodara. Her research interest includes Big Data, Cloud computing, Artificial Intelligence, data analytics and use of IT in business application. She has published paper in International journal and also presented paper in various national & international conferences. She is also an active member of various professional bodies like CSI, IETE and IE.



Prof. Dr. Jatinderkumar R. Saini completed Phd and MCA with Gold Medals in all three years from VNSGU, Surat. He is working as Professor and I/C Director at GNFC's Narmada College of Computer Application, Bharuch. He has served or is serving as executive committee member of Vadodara chapter of ISG, CSI and IETE.