

An Overview of Automatic Text Summarization Techniques

Sanjan S Malagi, Rachana
Radhakrishnan, Monisha R, Keerthana S
Department of Information Science &
Engineering,
JSS academy of technical education,
Bangalore-560060

Dr D V Ashoka
Professor, Department of Information
Science &
Engineering,
JSS academy of technical education,
Bangalore-560060

Abstract - In this age, where an enormous amount of information is available on the Web, it is most basic to supply the advanced mechanism to remove the data quickly and most profitably. It is uncommonly troublesome for human beings to physically extricate the outline of expansive records of content. There is a huge number of textual content, image content available on the Internet. So there's an issue, looking for significant documents from the number of reports accessible, and retaining significant data from it. To solve these issues, the programmed content summarization is exceptionally much necessary. Text Summarization (TS) is the method of recognizing the foremost imperative important data in a report or set of related reports and abstracting them into a shorter form protecting its general implications or meanings of the sentences. It is in this way we are ready to summarize the substance so that it gets simpler to ingest the information, keeping up the substance, and understanding the information. A few content summarization approaches have been presented in the past for a long time for English and some other languages. The fundamental objective is to decrease a given body of content to a division of its estimate, keeping up coherence and semantics. In this paper, we present a brief study of the various methods in existence to achieve this summarization.

Keywords - Text Summarization, imperative, coherence, languages, semantics.

I. INTRODUCTION

The web consists of different news articles, numerous images, blogs, and documents that are unstructured. The most feasible way to explore is to utilize and search and skim results. There may be an incredible need to decrease much of this content data to shorter, centered summaries that capture the notable points of interest, both so that we can explore it more viably as well as check whether the bigger reports contain the data that we are searching for. It moreover has applications in a few regions such as news artifacts, emails, web journal posts, research papers, and content reports to get a summary of the outcomes obtained. To discover important information, a client must go through the total records, this causes information over-burden issues which lead to a waste of time and pointless difficult work.

The most feasible solution of TS must consist of the following features which make the text summarization much easier for the user:

i) Reading the complete article, dismembering it, and isolating the vital thoughts from the raw content takes time

and needs hard work. Reading a commentary of five hundred words can take the slightest of 15 minutes. Programmed summary software summarizes writings of five hundred to five thousand words in seconds. The clients can get the foremost vital data to make strong conclusions despite having to study less information. Computers these days are far more capable than the human – and it is most likely that programs will make great summaries quickly before humans will have time to see into the article.

ii) A few summarization processes work in any dialect – a capacity that surpasses the capacities of most people. Since summarizers work on linguistic models can summarize writings in most dialects – from Western to European – without the requirement of human intervention to change the data. This makes them perfect for individuals who studied and dealt with multi-lingual information, or for individuals who got to change their data but wish to keep them as brief as conceivable.

iii) A special feature that the programs should have is the capacity to include the list of important original keywords in its summary such as 'intelligent', 'amazing' etc. which does not affect the original meaning of the data.

iv) Summarization techniques not only should summarize the text documents, but also should give out the summaries of the news articles directly from the web pages. This exceedingly improves efficiency because it speeds up the process of surfing. Instead of going through full news articles that contain large amounts of unwanted data, the user needs to only examine the outlines of such web pages that have an accurate and exact summary - but still retains 20% of the original article.

II. TEXT SUMMARIZATION METHODS

The earliest research of TS began in 1958, including the recurrence of a specific word, positioning of the sentence, and the presence of keywords. Since then, technology has come a long way and has consequently provided huge improvements in the field of automatic summarization. There are now multiple methods used to achieve accurate summaries. Moreover, the algorithms used also have applications in Search Engine Optimization. The major classification of text summarization methods is extractive and

abstractive. In this summary, we discuss some of these approaches used to achieve coherent and meaningful summaries.

A. Text Rank Algorithm

Automatic Text Summarizer was a paper written by Dr. Annapurna P Patilet al [1]. The authors describe three main processes in the Text-rank algorithm - i) Preprocessing: Text is separated based on punctuations, stop words, nouns, exclamation marks, blank spaces. These words are stemmed into the base form i.e., “continuing” or “continuous” etc. will be converted to its base form as “continue” and are listed. The final result consists of the words in their base form. The remaining words are stored in a list along with the corresponding sentences. ii) TextRank algorithm is implemented where each sentence (a hub within the chart) is compared with each other sentence. The match is allocated with a likeness score. This gives the similarity between the different occurring words and sentences and along with the occurrence of a number of repeated words. The given method is run until the weighted score of each hub does not alter the past 5 decimal places. Thereafter a weighted score for every sentence within the record is obtained. The sentences with the most elevated weights are chosen and are shown as per their introductory arrangement within the document. iii) The disintegration of the sentence as separate words with whitespaces while maintaining the fundamental sense of each word within the context associated with the given verb. In this case, a different word can be chosen and replaced which is diverse from the original word. The text-rank algorithm has shown a good result in most of the summarization methods. The major advantage of this method is that it works algorithmically without requiring advanced procedures. However, the replacement isn't adequately perfect as Natural language processing techniques are not used.

B. Term Frequency – Inverse Document Frequency

Term frequency (TF) and the inverse document frequency (IDF) are numerical insights that present how imperative a word is in a given report. TF can be referred to repetitive occurrence of the term in the report and IDF may be a measure that decreases the weight of repetitive terms within the collection and increases the weight of terms that are found rarely. At this point, sentences are scored concurring to items, and sentences with a high score are included in a rundown. One issue with this strategy is longer sentences often get high because they contain more number of words.

Categorized Text Document Summarization in the Kannada Language by Sentence Ranking was a paper written by Jayashree R et al [2]. The authors have discussed the use of a combination of GSS (Galavotti, Sebastiani, Simi) coefficients and IDF technique along with TF for the extraction, and summarizing documents in Kannada. Datasets contain documents obtained from Kannada Webdunia. The authors use a sentence scoring approach in which words and sentences are assigned scores and the top sentences are extracted. Finally, the authors compare the machine-generated

summary with a human-written summary. It can see that the summaries composed by people having information about the subject already are more exact than the summary produced by the machine. However, a machine-generated summary can be improved if techniques to achieve coherence are included. The authors believe that using Artificial Neural Network techniques can help solve this issue.

C. Natural Language Processing

NLP is a field of artificial intelligence that allows the computers to read, understand, and determine meaning from a human dialect in a keen and valuable way. It acts as a middle point of computer science, artificial intelligence, and computational etymology. Owing to the complicated nature of natural human language, processing of the data using this technique can prove to be difficult. Despite this complication, NLP has found a wide range of applications in different domains. By utilizing NLP, designers can organize and structure information to perform errands such as programmed summarization, interpretation, named substance acknowledgment, relationship extraction, sentiment analysis, speech recognition, and topic segmentation.

Automatic Text Summarization of News Articles is a paper written by Prakhar Sethi et al [3]. The authors have modified already existing methods to work with news articles only. The authors found that news articles are written in a certain pattern. They made use of this to refine and improve already existing methods to work for news articles with higher efficiency. The methods used in the proposed system were - A. Sentence Tokenization B. Part of speech tagging for tokenized words. C. Pronoun Resolution. D. Lexical Chain formation. E. Scoring Mechanisms - Lexical Chain Scoring, Sentence Scoring, Proper Noun Scoring. F. Summary Extraction - Extraction Based on Article Category, Using Sentence Scoring, Using Strong Lexical Chains, Using Proper Noun Scoring. The algorithm was tested on multiple inputs of varying lengths to obtain a better understanding of the pros and cons of the methods used. The authors were able to automatically summarize news articles and compare them and analyze which scoring parameter yields superior results.

A Survey of Automatic TS Techniques for Indian and Foreign Languages was a paper written by Prachi Shah et al [4]. In this paper, the authors explore techniques to work with various languages - Indian and Foreign. The three major steps proposed are

1. Preprocessing - segmentation, tokenization, and stop word removal is a part of the first phase of preprocessing.
2. Processing- some of the features considered in this step are sentence positioning, sentence length, numerical data, presence of inverted comma, and keywords in the sentence. Finally, a machine learning model is implemented to decide

if a particular sentence should be appended to the final summary or not.

3. Extraction - ranks are assigned to sentences based on their sentence scores. Higher ranked sentences are included in the summary.

Question and Answer Extraction Using NLP [14] written by Akash Shekar, Jeevanantham.P proposes a model to extract questions and answers given a specific topic. The authors have created this website using Python, HTML, PHP, Java, and MySQL database. The proposed site is utilized to extract information almost a given subject or word from Wikipedia and creates questions and answers for the given topic. The challenge is to create utilize of ML and NLP for address and reply era from a subject or paragraph. Systematic audits require master analysts to physically screen citations in arrange to distinguish all relevant articles. Our proposed strategy employments an address extraction based on the neural arrange vector space model. With the section vector strategy -find idle themes & clarifies substance of each point by meaningful and comprehensive content labels. In the existing framework, the NLP is utilized to look for as it were watchwords and characters but within the current framework, it creates questions and answers for given themes.

D. Optical Character Recognition

Optical Character Recognition (OCR) is the conversion of images containing typed text, manually written text, or printed/scanned content into machine-encoded material. It is the technology that extracts content from pictures so that a computer can perform actions on the extracted content.

Implementation of OCR for the Bengali Language was a paper written by Muhammed Tawfiq Chowdhury et al [5]. The authors consider two sets of inputs containing images converted from scanned images of printed documents. Detection of character and word are the two steps the proposed model works in. OCR also works on sentence detection. This aids in the maintenance of the structure of the document. The system works using the Tesseract OCR Engine with the UI developed in the Java Graphical User Interface platform. Unique identification of each and every character in the input file must be possible to enable OCR recognition. The foremost drawback is that spaces among words are not recognized by OCR which diminishes the significance of the yield to a couple of extents.

E. K-Nearest Neighbor and Naïve Bayes Classifier

KNN is an apathetic learning, non-parametric calculation. It uses information with a few classes to predict the categorization of the unused test point. KNN is non-parametric since it does not make any presumptions on the information being considered. K Nearest Neighbor may be a basic calculation that stores all the accessible cases and classifies the unused information or case based on a closeness degree.¶

Naive Bayes may be a basic, however efficient and commonly-used, ML classifier. It may be a probabilistic classifier that creates classifications utilizing the Maximum A Posteriori choice rule in a Bayesian setting. It can be portrayed using an exceptionally simple Bayesian network. The Naive Bayes algorithm is a supervised ML algorithm which is mainly used for classification purposes. The Naive Bayes calculation is called “naive” since it makes the presumption that the event of a certain characteristic is autonomous of the occurrence of other characteristics. Naive Bayes classifiers have been particularly prevalent for content classification. In this strategy, the training information set is utilized for reference to create a rundown. The summarization process is modeled as a classification issue. Sentences are classified based on the highlights that they have.

Feature Selection Methods for Classifying Email Messages [6]: Analysis, Proposal, and Comparative Study was a paper written by Sanaa Abou Elhamayed et al. The authors discuss the classification of emails to differentiate spam and non-spam emails. KNN and Naive Bayes classifiers are used to achieving this. Additionally, feature selection methods help in the selection of significant highlights in the emails, and others are discarded to get rid of processing overheads. By making a comparison of many such techniques, the authors conclude that the applications of this can span to other domains also. The authors observed variation inaccuracy in changing the number of selected features.

Automatic TS based on sentence clustering and extraction was a paper written by Zhang Pei-Ying et al [7]. This paper proposes a sentence similarity computing method based on the three features of the sentences - analyzing the word form feature, the word order feature, and the semantic feature. The approach consists of three steps:

1. Clustering of the sentences depending on the semantic distance.
2. Each cluster calculates the total sentence likeness (multi-features combination method).
3. Subject sentences are chosen by a few extraction rules. Weights are utilized to portray the commitment of each sentence feature.

The proposed framework employs the weight, to depict the commitment of each including the sentence, portraying the sentence similarity more precisely.

F. Semantic Analysis

Automatic Text Summarization Based on Semantic Analysis Approach for Documents in the Indonesian Language written by PanduPrakoso Tardan [8]. The paper discusses the main concept behind the semantic analysis is to obtain the sentence similarities by the vector value calculation of each sentence along with the title. The semantic analysis starts with the relationship between words. WordNet has been used to identify the depth of each word for their word similarity. This paper contains an approach to Indonesian language analysis.

The author has proposed a system having 4 main phases in the summarization:

- (1) Pre-processing for the elimination of wastes and unimportant words existing in the original document. This phase also involves the extraction of sentences, tokenization, removal of stop words, N-Gram detection, and stemming.
- (2) Feature computation phase which aims at the extraction and the computation of the sentence similarity values amongst sentences along with their title.
- (3) Feature ranking aims at the arrangement of the sentences in descending order based on their rankings. i.e., from higher rank to lower rank.
- (4) Once this phase is completed, a readable summary is generated that has extracted a gist of information from the actual document based on their rankings.

G. Artificial Intelligence

AI is a technique to form a computer, a robot, or an item to think about how smart humans think. AI may be a consideration of how the human brain thinks, learns, chooses, and works when it tries to illuminate issues. And at last, this study yields brilliant computer program systems. The point of AI is to progress computer capacities which are related to human information, for example, thinking, learning, and problem-solving. Artificial neural networks (ANN) are an important machine learning technique. ANN can be depicted as a mathematical model that is inspired by the way biological neural networks within the human brain handle data. A neural network is prepared to memorize the important highlights of sentences that should be included within the outline of the article. The neural network at that point is adjusted to generalize and combine the significant highlights clear in outline sentences.

AI for Automatic Text Summarization was a paper written by Min-Yuh Day et al [9]. In this paper, the author proposed how AI is used for Automatic TS by applying deep learning to produce short summaries from the large data. The author discusses the main objectives like the use of AI technologies which consists of statistical methods, machine learning, and deep learning, to create candidate titles, and compare the exactness.

Deep learning is a branch of machine learning, and it is used to do a multi-linear calculation and automatically create highlights with tremendous information and more than one covered up layer. Automatic text Summarization using deep learning consists of following steps like

- 1) Obtain data - This step includes getting the raw data from the web source.
- 2) Data preprocessing - This phase consists of two steps that are flitting special characters and convert encode.
- 3) Model development - In this step, the data is included in the three distinct models to produce three varieties of candidate titles.
- 4) Evaluation - This step uses ROUGE as an evaluation method to give the value of the candidate title.

This paper discusses the application of deep learning to produce simple outlines, comparing with distinct methods and training and testing of English exposition titles and abstracts from 1970 to 2017.

H. Text/Image Mining

Text mining is commonly referred to as information extraction or data mining. Text mining can be defined as the derivation of the valuable part of the data from the given text or text document. This important information is usually derived through identifying trends and patterns through a method called statistical pattern learning.

A Survey of Automatic Text Summarization Techniques for Indian and Foreign Languages Prachi Shah et al [10]. The authors have investigated innumerable research projects and found that there are various techniques of automatic TS systems for languages like English, European languages, and Asian languages. It has been noticed that there is comparatively less research work done for Indian languages. Based on this analysis here, an automatic Hindi TS using an ML technique was proposed. Machine Learning can be refunded as a technique which combines various indicators and decides the kind of features to be used and the method of information extraction. In this approach, the system has 3 major blocks: Pre-processing stage that involves segmentation, tokenization, and removal of stop words. The processing stage is the important step in the summarization involving feature extraction, sentence positioning, numerical data extraction, frequency of the words selection, presence of special characters, and keywords in the sentence. Extraction being the final phase, the ML method is used for the identification of the importance of the sentence based on the training set. The sentence is then given a rank based on the sentence score. Only the top-ranked sentences are used in the summary.

Content-Based Image Retrieval (CBIR) Public and Private Search Engines [13] by Dr. T. Santha, M. Abhayadev discusses some techniques for image mining based on

content. Due to the increase in the amount of digital data and images on the internet every day, this method discusses retrieval techniques that would help save time. Content-Based Visual Information Recovery (CBVIR) is the application of computer vision procedures to the image retrieval issue, that is, the issue of looking for digital pictures in expansive databases. CBIR is contradicted by conventional concept-based approaches. CBIR picture recovery preparation will offer assistance for overcoming the picture mining issues. Google, Yandex, and other look motors are utilized for picture retrieval purposes. CBIR innovation is much superior to old techniques.

I. Fuzzy Logic

Fuzzy logic is used to estimate the degree of significance and relationship and also to highlight the vital phrases to form summarization. Fuzzy Logic is mainly based on natural language. It may be a numerical device utilized for dealing with uncertainty, imprecision, ambiguity, and unclearness. Fuzzy logic could be a form of many-valued logic that bargains with inexact reasoning instead of settled and exact reasoning. Fuzzy logic is utilized to handle the concept of halfway truth where its truth esteem ranges between fully genuine and fully wrong.

Automatic Text Summarization using Fuzzy Inference is a paper written by Mehdi Jafari et al [11]. In the proposed model the authors coordinated the fuzzy logic with traditional extractive and abstractive approaches for content summarization. Through such a procedure, we get an outline of the original text, which can review the most conveyance of the original context. The overall flowchart of text summarizer proposed in the paper consists of the following phases

1) *Preprocessing phase* - Within the preprocessing stage, to convert the initial content into the text that will be utilized as an input parameter within the summarization framework, the following steps should be taken like Removing the redundant words: The words that do not have any value and contain no specific data like “an”, “the” etc. are removed. Case Folding: Either capitalized letters are transformed into lowercase or all lowercase are transformed into uppercase. Here we change overall characters into the lowercase. Stemming: The inferred words are transformed into their stemming. After this preprocessing phase, the next phase is the extraction of Syntactic parameters.

2) *Extraction of syntactic parameters* - This phase consists of Sentence Length (Finding the most important words), Location of Sentence(Finding the position of the important words in the sentences), Similarity to the title (Calculating the similarity between the title and the sentence), Similarity to the keywords (Calculating the similarity between the keywords), Text-to-text Coherence, Integrated text-to-center.

3) *Extraction of the semantic parameters* - This phase includes Semantic similarities between sentences (finding the

similarities between two sentences) and the order of words (find the order of the words in the sentences).

4) *Calculating sentence grades based on fuzzy logic* - This phase includes grading the sentences based on fuzzy logic. The fuzzy logic is utilized for measuring the degree of significance and relationship and also used to distinguish the vital sentence to create summarization.

5) *Summary Production* - This phase consists of a combination of parameters from the previous phases to produce an efficient summary.

III. CASE STUDY

In this case study [12], the Automatic Text Summarization is done using Natural Language Processing technique for preprocessing textual data, Optical Character Recognition achieved using Tesseract OCR to preprocess image data and Text-Rank algorithm to model the content and generate a summary. This system works based on accepting 4 types of input data from the user and performs processing and summarization methods on the input data and gives an extractive and coherent summary for the user based on their requirement. This generated summary can be converted into an mp3 file format for audio output (using the text-to-speech concept) and well as be downloaded from the site. The generated summary can also be translated to another language from a given list of languages using google translate.

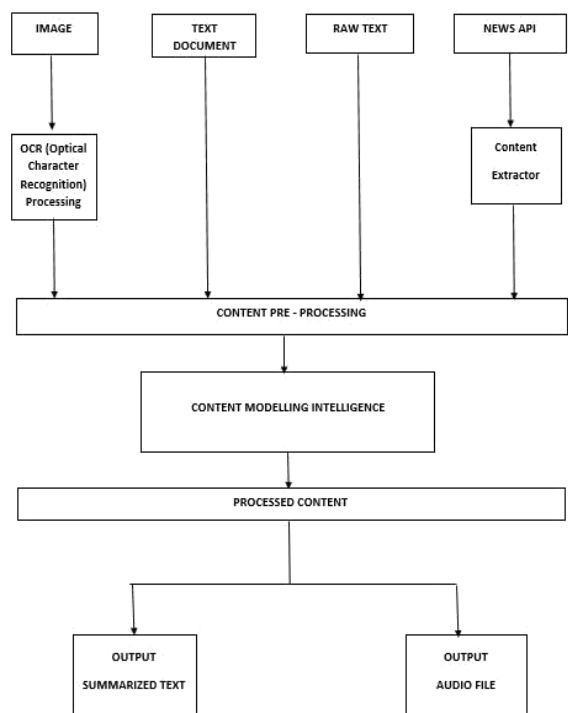


Fig. 1 Methodology used in Automatic Text Summarization. [12]

It includes 4 different inputs: Image processing, Text document processing, Raw text processing, and News article processing as shown in Fig 1. The image data is taken out by using Optical Character Recognition and preprocessing of data is done in the Content Modelling Intelligence [12] and the processed content is organized in proper meaningful sentences as output in both text and audio format.

IV. OBSERVATIONS AND RESULTS

From the above-discussed techniques, we find that a lot of the mentioned methods show promises in providing accurate and coherent summaries and the following points have been observed:

- The main job before the extractive summarization is to sight the required data that is to be summarized.
- The accuracy of the summary is largely dependent on the accuracy of the provided input data.
- The extracted sentences usually are longer than average since it contains unnecessary information in the summary.
- Redundant information might get included in the summary.
- Extraction based summaries are not very convenient to go through and process.
- There is a lack of flow of information in the extracted summary since each part of the contents is selected from different parts of the text file which sometimes leads to topic shift.
- Abstractive summary sometimes might not denote the semantic relationship between the important terms of the document.
- NLP- Natural Language Processing is very necessary for generating a meaningful summary.
- The abstractive summary quality might be low because of the lack of understanding of the semantic relationship between the words and the linguistic skills.

V. CONCLUSION

This survey paper covers different types of summarization processes based on extractive and abstractive techniques by using different algorithms for the summarization. Summarization processes have to produce a compelling summary in a brief time with less redundancy having linguistically correct sentences. All the above techniques used to give out good outcomes and also efficient summaries are obtained according to the context used. However, the challenge arises in creating accurate summaries with proper semantics including all the features of text summarization, image summarization, article summaries, and multiple document summary. The case study discussed also integrates translation and text-to-speech features that make the model more versatile and makes the model more user friendly. It can

be observed that a combination of the preprocessing and processing techniques discussed above could give an efficient model containing all relevant features to the model user friendly.

REFERENCES

- [1] Dr. Annapurna P Patil, Shivam Dalmia, Syed Abu Ayub Ansari, Tanay Aul, Varun Bhatnagar, "Automatic Text Summarizer", *International Conference on Advances in Computing, Communications and Informatics, IEEE* (2014).
- [2] Jayashree R, Shreekantha Murthy K, "Categorized Text Document Summarization in the Kannada Language by Sentence Ranking", *12th International Conference on Intelligent Systems Design and Applications (ISDA), IEEE* (2012).
- [3] Prakhar Sethi, Sameer Sonawane, Saumitra Khanwalker, R. B. Keskar, "Automatic Text Summarization of News Articles", *International Conference on Big Data, IoT and Data Science (BIG) Vishwakarma Institute of Technology, Pune, Dec 20-22 IEEE* (2017).
- [4] Prachi Shah, Nikitha P. Desai, "A Survey of Automatic Text Summarization Techniques for Indian and Foreign Languages", *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)* (2016).
- [5] Muhammed Tawfiq Chowdhury, Md. Saiful Islam, Baijed Hossain Bipul and Md. Khalilur Rhaman, "Implementation of an Optical Character Reader (OCR) for Bengali Language", *International Conference on Data and Software Engineering IEEE* (2015).
- [6] Sanaa Abou Elhamayed, Samah Osama M. Kamel, "Feature Selection Methods for Classifying Email Messages: Analysis, Proposal, and Comparative Study", *Int. J. Advanced Networking and Applications* (2019).
- [7] ZHANG Pei-Ying, LI Cun-he, "Automatic text summarization based on sentence clustering and extraction", *Institute of Electrical and Electronics Engineers IEEE* (2009).
- [8] Pandu Prakoso Tardan, Alva Erwin, Kho I Eng, Wahyu Muliady, "Automatic Text Summarization Based on Semantic Analysis Approach for Documents in Indonesian Language", *978-1-4799-0425-9/13/\$31.00 ©2013 IEEE* (2013).
- [9] Min-Yuh Day, Chao Yu Chen, "Artificial Intelligence for Automatic Text Summarization", *International Conference on Information Reuse and Integration for Data Science IEEE* (2018).
- [10] Prachi Shah, Nikhita P Desai, "A Survey of Automatic Text Summarization Techniques for Indian and Foreign Languages", *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)* (2016).
- [11] Mehdi Jafari, Jing Wang, Yongrui Qin, Mehdi Gheisari, Amir Shahab Shahabi, Xiaohui Tao, "Automatic Text Summarization Using Fuzzy Inference", *IEEE* (2013).
- [12] Sanjan S Malagi, Rachana Radhakrishnan, Monisha R, Keerthana S, Dr D. V. Ashoka, "Content Modelling Intelligence System Based on Automatic Text Summarization", *International Journal of Advanced Networking and Applications Volume:1 Issue:1 Pages:1-9 (2020) ISSN: 0975-0290 (IJANA)* (2020).
- [13] Dr T. Santha, M. Abhayadev "Content Based Image Retrieval Public and Private Search Engines", *Special Issue Published in Int. Jnl. Of Advanced Networking and Applications (IJANA) pp:98-102(2015)*
- [14] Akash Shekar, Jeevanantham.P, "Question and Answer Extraction Using NLP", *Special Issue Published in Int. Jnl. Of Advanced Networking and Applications (IJANA)197-199*
- [15] Dr. D V Ashoka, Sanjay B Ankali, "Detection Architecture of Application Layer DDoS Attack for Internet", *International Journal of Advanced Networking and Applications* (2011).
- [16] R. Chetan and D. V. Ashoka, "Data mining-based network intrusion detection system: A database centric approach", *2012 International Conference on Computer Communication and Informatics, Coimbatore, pp. 1-6, doi: 10.1109/ICCCI.2012.6158816* (2012).

BIBLIOGRAHIES

Sanjan S Malagi, Rachana Radhakrishnan, Monisha R, and Keerthana S is presently studying in BE final year Information Science & Engineering, JSS ACADEMY OF TECHNICAL EDUCATION, Bangalore. Their area of interest includes Cloud Computing, Internet of Things (IOT), Natural Language Processing, Text Mining, AI and Bigdata.

Dr D. V. Ashoka is presently working as Professor in the Department of Information Science and Engineering at JSS Academy of Technical Education, Bangalore. He received his MTech. in computer science & Engineering from VTU and Ph.D. degree in Computer Science and Engineering from Dr. MGR, University, Chennai. He has 25 years of academic and research experience. He has published more than 60 research papers in national / international journals and conferences. His fields of interest are Requirement Engineering, Operating System, Computer Organization, Software Architecture, data mining in image processing and Cloud Computing.