**Text Summarization**

# Abstract:

The area of ​​Natural Language Processing (NLP) is a subarea of ​​Artificial Intelligence that aims to make computers capable of understanding human language, both written and spoken. Some examples of practical applications are: translators between languages, translation from text to speech or speech to text, chatbots, automatic question and answer systems (Q&A), automatic generation of descriptions for images, generation of subtitles in videos, classification of sentiments in sentences, among many others! Another important application is the automatic document summarization, which consists of generating text summaries. Suppose you need to read an article with 50 pages, however, you do not have enough time to read the full text. In that case, you can use a summary algorithm to generate a summary of this article. The size of this summary can be adjusted: you can transform 50 pages into only 20 pages that contain only the most important parts of the text!

# Literature review:

Chin-Yew Lin [1] In this paper the author introduced Recall Oriented Understudy for Gusting Evaluation ROUGE. That is an automatic evaluation package for text summarization. The paper also introduced four different measures of ROUGE: - ROUGE-N, ROUGE-L, ROUGE-W and ROUGE-S. It measures the quality of summary by comparing the generated summary with other ideal summaries that are created by humans. These methods are efficient for automatic evaluation of single document summary as well as multi-document summaries.

Akshil Kumar et al. [2] In this paper the author has analyzed and compared the performance of three different algorithms. Firstly, the different text summarization techniques explained. Extraction based techniques are used to extract important keywords to be included in the summary. For comparison three comparison three keyword extraction algorithms namely TextRank, LexRank, Latent Semantic Analysis (LSA) were used. Three algorithms are explained and implemented in the python language. The ROUGE 1 is used to evaluate the effectiveness of the extracted keywords. The results of the algorithms are compared with the handwritten summaries and evaluate the performance. In the end, the TextRank Algorithm gives a better result than the other two algorithms.

Pankaj Gupta et al. [3] In this paper the author has reviewed different techniques of Sentiment analysis and different techniques of text summarization. Sentiment analysis is a machine learning approach in which a machine learns and analyzes the sentiments, emotions present in the text. Machine learning methods like Naive Bayes Classifier and Support Machine Vectors (SVM) are used. These methods are used to determine the emotions and sentiments in the text data like reviews about movies or products. Text summarization uses the natural language processing (NPL) and linguistic features of sentences to check the importance of the words and sentences that can be included in the final summary. In this paper, a survey has been done of previous research work related to text summarization and Sentiment analysis, so that new research areas can be explored by considering the merits and demerits of the current techniques and strategies.

Harsha Dave et al. [4] In this paper the author has proposed a system to generate the abstract summary from the extractive summary using WordNet ontology. The multiple documents had been used like text, pdf, word files etc. The author has discussed various text summarization techniques then the author discussed step by step the multiple document text summarization approaches. The experiment result is compared with the existing online extractive tools as well as with human generated summaries and shows the proposed system gives good results. At last the author proposed for the future that the summarization accuracy can be increased by comparing this abstractive system with some other abstractive system.

Yihong Gong et al. [5] In this research paper the author proposes two methods that create the generic text summaries by ranking and extracting sentences from the main text documents. The first method uses information retrieval (IR) methods that rank the sentence relevance and provides the relevance scores to sentences and the second method uses the latent semantic analysis (LSA) technique that based on latent semantic indexing (LSI) in order to identify the semantic importance of the sentences, for summary creations. The author uses the Singular Value Decomposition (SVD) to generate the text summary. Further, this paper author explains the SVD based methods step by step. The effect of different Weighted Schemes is also checked on the performance of the summaries. The proposed methods provide generic abstractive summaries. Finally, the results are compared with the human-generated summaries. It generates better human-like abstractive summaries. The future author proposed to investigate various machine learning techniques so that the quality of generic text summarization can be improved.

Dharmendra Hinhu et al. [6] In this paper the author uses the extractive text summarization. The author gives the Wikipedia Articles as input to the system and identifies text scoring. Firstly, the sentences are Tokenized through pattern matching using regular expressions. Then we get data in the form of a set of words then stop words are removed from the set of words. The words are then stemmed. Then traditional methods are used for scoring of the sentences. Scoring helps in classifying the sentences if they are included in summary or not. It is found that scoring sentences based on citation gives better results.

Rada Mihalcea et al. [7] In this paper the author introduced the TextRank, a graph-based ranking model for the processing of the text. It is an unsupervised method for keyword and sentence extraction. TextRank uses voting based weighting mechanism and provides the score to the sentence then finally determines the importance of the sentence. The nodes in the graph represent the sentences. The significance of the sentence based on incoming and outgoing edges from nodes. The weight of each is determined based on the similarity score between the sentences. TextRank derived from the Google’s Page Rank algorithm. TextRank provides extractive summaries of the text. Text Rank Provides the best results.

N. Moratanch et al. [8] In this paper the author presents an exhaustive survey on abstraction-based text summarization techniques. The paper presents a survey on two broad abstractive summary approaches: Structured based abstractive summarization and Semantic-based abstractive summarization. The author presents the review of various researches on both approaches of abstractive summarization. The author also covered the various methodologies and challenges, in abstractive summarization.

# Methodology:

1. Pre-processing the texts
2. Find Word frequency
3. Find Weighted word frequency
4. Sentence tokenization
5. Score for the sentences
6. Order the sentences
7. Generate the summary

# Results and discussion:

## Original Text:

Artificial intelligence (AI) is intelligence demonstrated by machines, as opposed to natural intelligence displayed by animals including humans. Leading AI textbooks define the field as the study of "intelligent agents": any system that perceives its environment and takes actions that maximize its chance of achieving its goals.[a] Some popular accounts use the term "artificial intelligence" to describe machines that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving", however, this definition is rejected by major AI researchers.

## Summarized Text:

Some popular accounts use the term "artificial intelligence" to describe machines that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving", however, this definition is rejected by major AI researchers. Leading AI textbooks define the field as the study of "intelligent agents": any system that perceives its environment and takes actions that maximize its chance of achieving its goals. Artificial intelligence (AI) is intelligence demonstrated by machines, as opposed to natural intelligence displayed by animals including humans.

# References:

1.  Chin-Yew Lin, “Rouge: A Package for Automatic Evaluation of Summaries.” Barcelona Spain, Workshop o Text Summarization Branches Out, Post- Conference Workshop of ACL 2004.

2.  Akshi Kumar, Aditi Sharma, Sidhant Sharma, Shashwat Kashyap, “Performance Analysis of Keyword Extraction Algorithms Assessing Extractive Text Summarization.” International Conference on Computer, Communication, and Electronics (Comptelix), 2017.

3.  Pankaj Gupta, RituTiwari and Nirmal Robert, “Sentiment Analysis and Text Summarization of Online Reviews: A Survey.” International Conference on Communication and Signal Processing, 2016.

4.  Harsha Dave, Shree Jaswal, “Multiple Text Document Summarization System using Hybrid Summarization Technique.” 1st International Conference on Next Generation Computing Technology (NGCT), 2015.

5.  Yihong Gong, Xin Liu, “Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis.” Proceeding of the 24th annual international ACM SIGIR Conference on research and development in information retrieval, ACM 2001.

6. DharmendraHingu, Deep Shah, Sandeep S.Udmale, “Automatic Text Summarization of Wikipedia Articles.” International Conference on Communication, Information & Computing Technology (ICCICT), 2015.

7. Radha Mihalcea, Paul Tarau, “TextRank: Bring Order into Texts.” Association for Computational Linguistics, 2004.

8. . N. Moratanch, Dr. S. Chitrakala, “A Surveyon Abstractive Text Summarization.” International Conference on Circuit, Powe and Computing Technologies (ICCPCT), 2016.