iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IX **Month of publication:** September 2022

DOI: https://doi.org/10.22214/ijraset.2022.46925

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IX Sep 2022- Available at www.ijraset.com

Automated Generation of Summary in Six Different Ways from Any Paragraph and Automated Generation of Fill in the Blanks from Any Paragraph

Rutvik Rajiv Durve
Somaiya Vidya vihar University

Abstract: ***Many students finds difficult to read the entire contents during the examination and hence fail to write in the exams too. However, it is required to build the system that can overcome their problems. This paper talks about how to do summary of text based on the Text Summarization algorithm and various different models. For example, the alphabet is the simplest structure, in that it is a collection of letters that can form strings called words. A formal language is one that has regular, context-free, and formal grammar. In addition to the development of computer sciences as a whole, artificial intelligence's advancements also played a role in our continuing understanding of NLP.***
Index Terms: Text Summarization, Luhn Algorithm, TF-IDF algorithm

I. INTRODUCTION

Our application is called Laconic. It's a standalone application that can run on operating systems like UNIX, MacOS, Google Chrome and Windows etc. The programming language used to build up the software is Python and Natural Language Processing, a branch of Machine Learning. Extensive use of user friendly interface.
Below is the function of Laconic function:

1) It can generate six types of summary like General summary, Summary by High Ranking words, Summary by keywords, Summary in simple words from any given paragraph.
2) It can also generate automated Fill in the blanks with questions and answers separately.
3) Users can easily copy and paste the text generated by summary and fill in the blanks with questions and answers.

II. LITERATURE SURVEY

Research work [03] proposed approach retrieves the important information from the text by performing semantic analysis of the text. In the approach Simplified Lesk Algorithm is used to extract the relevant sentences from a single-document text based on the semantic information of the sentence and WordNet is used as an online semantic dictionary
Research work [07] proposes a sentence similarity computing method based on the three features of the sentences, on the base of analyzing of the word form feature, the word order feature and the semantic feature, using the weight to describe the contribution of each feature of the sentence, describes the sentence similarity more preciously.
[14] reviews text summarization approaches and recent deep learning models for this approach. Additionally, it focuses on existing datasets for these approaches, which are also reviewed, along with their characteristics and limitations. The most often used metrics for summarization quality evaluation are ROUGE1, ROUGE2, ROUGE L, and Bleu.
[15] gives the comparison of various neural network based abstractive text summarization models and also discuss the types of summarization based on categories and different approaches of abstractive text summarization.
Research work [16] outlines extractive and abstractive text summarization technologies and provides a deep taxonomy of the ATS domain. The taxonomy presents the classical ATS algorithms to modern deep learning ATS architectures.
[18] is a survey on the various types of text summarization techniques starting from the basic to the advanced techniques. According to this survey, seq2seq model along with the LSTM and attention mechanism is used for increased accuracy.
[23] In this study, we review and compare three extractive summarization methods that are TextRank, Sentence Scoring and conceptual data classification method for summarization.

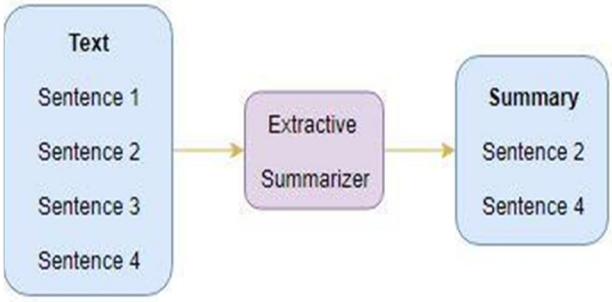
©IJRASET: All Rights are Reserved

From the literature survey done, it has been observed that a summarization process has been presented characterized by the inclusion of words recognition algorithm, paragraph clustering algorithm and summary generation algorithm. Automatic text summarization is a complex task which contains many sub-tasks in it. Every subtask has an ability to get good quality summaries.
There are a lot of different techniques that can be usedfor Text Summarization such as Gensim, Tensor flow, and Transformers. It has been observed that The hardest NLP tasks are the ones where the output isn't a single label or value (like Classification and Regression), but a full new text (like Translation, Summarization and Conversation)
From the literature review done, it seems a BoW model is one of the more simplistic feature extraction algorithms that you will come across. The name "bag-of-words" comes from the algorithm simply seeking to know the number of times a given word is present within a body of text. The order or context of the words is not analyzed here. Similarly, if we have a bag filled with six pencils, eight pens, and four notebooks, the algorithm merely cares about recording the number of each of these objects, not the order in which they are found, or their orientation.

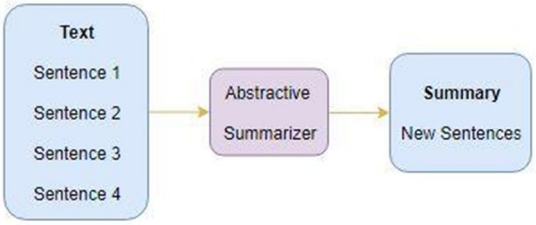
III. METHODOLOGY

One application of text analytics and NLP is Text Summarization. Text Summarization Python helps in summarizing and shortening the text in the user feedback. It can be done with the help of an algorithm that can help in reducing the text bodies while keeping their original meaning intact or by giving insights into their original text.
Two different approaches are used for Text Summarization:

1) *Extractive Summarization:* In Extractive Summarization, we are identifying important phrases or sentences from the original text and extract only these phrases from the text. These extracted sentences would be the summary.



2) *Abstractive Summarization:* In the Abstractive Summarization approach, we work on generating new sentences from the original text. The abstractive method is in contrast to the approach that was described above. The sentences generated through this approach might not even be present in the original text.



We are going to focus on using extractive methods. This method functions by identifying important sentences or excerpts from the text and reproducing them as part of the summary. In this approach, no new text is generated, only the existing text is used in the summarization process.
Rake algorithm:
We have use Rake() in our application. RAKE is that keywords frequently contain multiple words but rarely contain punctuation, stop words, or other words with minimum lexical meaning.
Once we have the text corpus, RAKE splits the text into a list of words, removing stop words from the same list. Return list is known as Content Words. Folks familiar with Natural language processing are aware of the terms stop words. Words like 'are', 'not', 'there,' 'is' do not add any meaning in a sentence. Ignoring them will make our main corpus lean and clean.

IV. RESULTS

Let's create a Word degree matrix like below, where each row will display the number of times a given content word co-occurs with another content word in candidate key phrases.

	feature	extraction	complex	algorithm	available	help	rapid	automatic	keyword
feature	2	2	0	0	0	0	0	0	0
extraction	2	3	0	0	0	0	0	1	1
complex	0	0	1	0	0	0	0	0	0
algorithm	0	0	0	1	1	0	0	0	0
available	0	0	0	1	1	0	0	0	0
help	0	0	0	0	0	1	0	0	0
rapid	0	1	0	0	0	0	1	1	1
automatic	0	1	0	0	0	0	1	1	1
keyword	0	1	0	0	0	0	1	1	1
TOTAL SUM	**4**	**8**	**1**	**2**	**2**	**1**	**3**	**4**	**4**

Candidate Key Phrases :
- ✓ feature extraction
- ✓ Complex
- ✓ help
- ✓ algorithms available
- ✓ rapid automatic keyword extraction

Total **Unique** candidate key phrases count is 5. but total 'Feature' is coming with 2 candidate key phrases . Hence 2

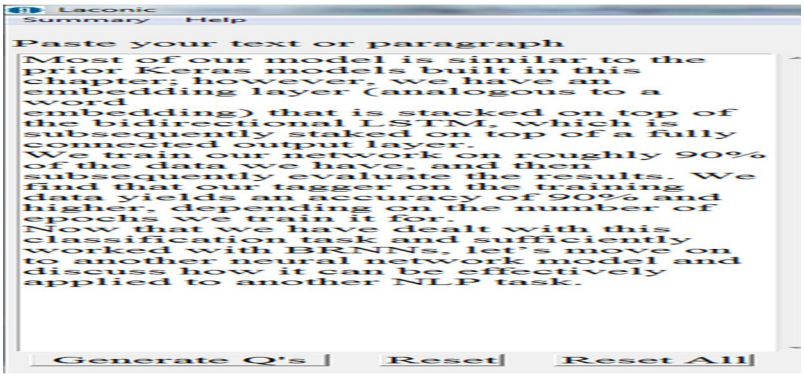
We have to give a score to each word. Calculate the 'degree of a word' in the matrix, the sum of the number of co-occurrences, then divide them by their occurrence frequency. Occurrence frequency means how many times the word occurs in the primary corpus.

Word	Degree Of word	Word Frequency	Degree Score
feature	4	2	2
extraction	8	3	2.66
complex	1	1	1
algorithm	2	1	2
available	2	1	2
help	1	1	1
rapid	3	1	3
automatic	4	1	4
keyword	4	1	4

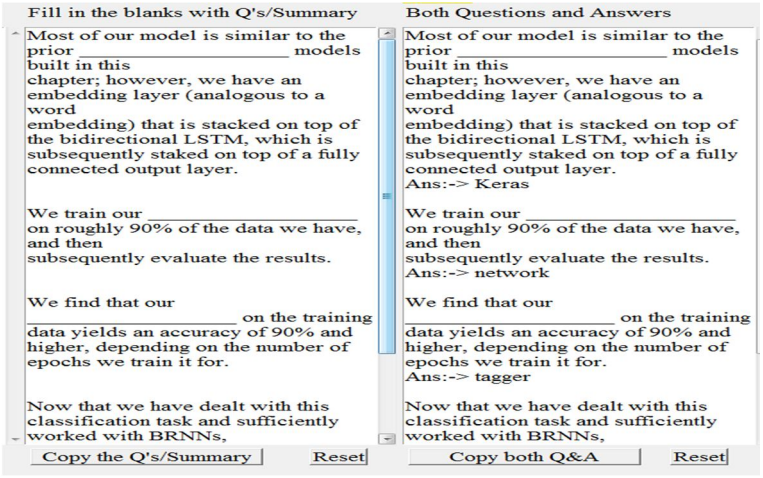
Now consider the candidate key phrases and consider a combined(sum) score of each word for each candidate key phrases.

Cumulative Score of candidate Key Phrases				
feature	extraction			4.66
Complex				1
algorithms	available			4
help				1
rapid	automatic	keyword	extraction	13.66

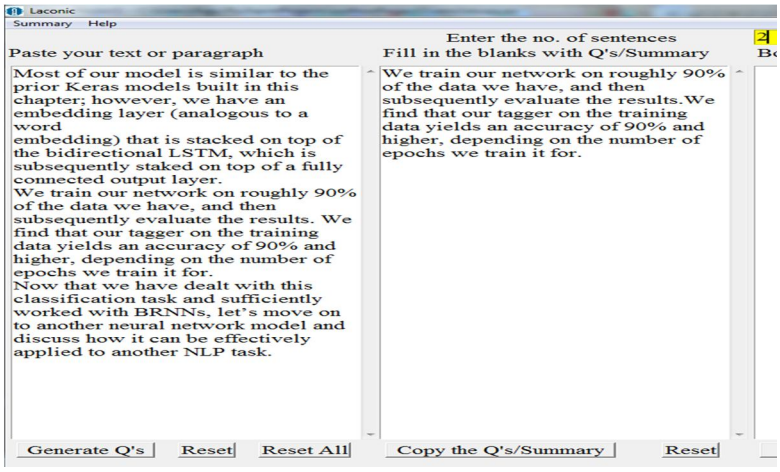
Suppose two keywords or key phrases appear together in the same order more than twice. A new key phrase is created regardless of how many stop words the key phrases contained in the original text. The score of the key phrases is computed just like the one for a single key phrase.



In the above image the user have added the text in the text area component. If the use clicks on Generate Q's button, it will give output in another Text area in the form of Fill in the blanks with questions and answers.



In the above pic user can see Fill in the blanks with questions and answers are generated by the application. If the user click on Copy the Q's/Summary button he can copy and paste on editor only questions and summary and if the user clicks on Copy both Q&A button he can copy and paste on editor both questions and answers.



The above result after processing gives the output using a summary by a number of sentences as 2 using keywords. In the First box, the user can place the paragraph. On top of the window, we have a summary menu that has a sub-menu like summary by simple words, summary by keywords, etc. If the user clicks on any sub-menu will get the result on the second text box.

V. CONCLUSION

The NLP community finds text summarizing to be intriguing study areas that can assist create information that is concise. The purpose of this paper is to describe the most recent findings and developments in this area using the Rake approach. The relationship between trends/topics, problems and challenges in each topic, technique, and method used is summarized into one to make it easier to explore and re-analyze.

In the topic of text summarization research, future work that can be done includes i) solving feature problems, namely determining the most appropriate features to use in summarizing by the dataset by selecting features, discovering new features, optimizing commonly used features, feature engineering, using features for semantics, linguistic features, finding features to produce coherence sentences, and add grammatical features. ii) Preprocessing by the problem dataset using appropriate stemming, besides, to stop word removal and tokenizing, POS Tagging is also needed, namely to categorize word classes, such as nouns, verbs, adjectives, etc. iii) For extractive summaries, collaborating statistical techniques, fuzzy-based techniques, and machine learning are very challenging to try.

VI. ACKNOWLEDGEMENT

We sincerely grateful to SK Somaiya University for providing support related open access in this paper. Moreover, we would like to express many thanks to HOD-(Head of Department of Computer Science) Mrs. Swati Maurya for providing insights and expertise that greatly assist in this research.

REFERENCES

[1] R. Pal and D. Saha, "An approach to automatic text summarization using WordNet," 2014 IEEE International Advance Computing Conference (IACC), 2014, pp. 1169-1173, doi: 10.1109/IAdCC.2014.6779492.

[2] P. -y. Zhang and C. -h. Li, "Automatic text summarization based on sentences clustering and extraction," 2009 2nd IEEE International Conference on Computer Science and Information Technology, 2009, pp. 167-170, doi: 10.1109/ICCSIT.2009.5234971.

[3] A. Elsaid, A. Mohammed, L. F. Ibrahim and M. M. Sakre, "A Comprehensive Review of Arabic Text Summarization," in IEEE Access, vol. 10, pp. 38012-38030, 2022, doi: 10.1109/ACCESS.2022.3163292.

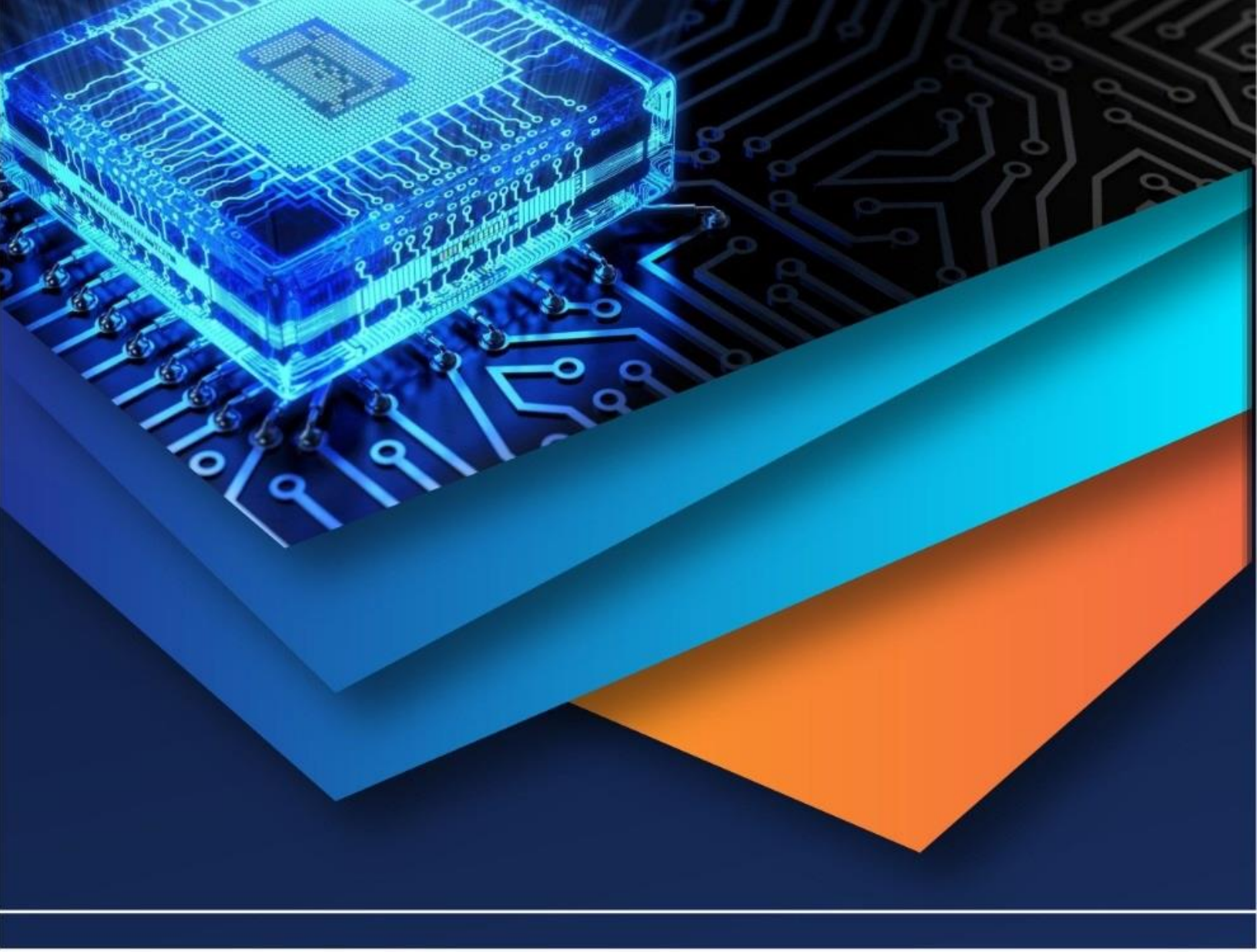
[4] J. Tandel, K. Mistree and P. Shah, "A Review on Neural network based Abstractive Text Summarization models," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019, pp. 1-4, doi: 10.1109/I2CT45611.2019.9033912.

[5] R. Boorugu and G. Ramesh, "A Survey on NLP based Text Summarization for Summarizing Product Reviews," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 352-356, doi: 10.1109/ICIRCA48905.2020.9183355.

[6] M. F. Mridha, A. A. Lima, K. Nur, S. C. Das, M. Hasan and M. M. Kabir, "A Survey of Automatic Text Summarization: Progress, Process and Challenges," in IEEE Access, vol. 9, pp. 156043-156070, 2021, doi: 10.1109/ACCESS.2021.3129786.

[7] A. W. Palliyali, M. A. Al-Khalifa, S. Farooq, J. Abinahed, A. Al-Ansari and A. Jaoua, "Comparative Study of Extractive Text Summarization Techniques," 2021 IEEE/ACS 18th International Conference on Computer Systems and Applications (AICCSA), 2021, pp. 1-6, doi: 10.1109/AICCSA53542.2021.9686867.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)