International Journal of Innovative Research in Engineering
Volume 3, Issue 3 (May-June 2022), PP: 396-398
www.theijire.com
ISSN No: 2582-8746

Text Summarization Using Word Frequencies

Punyaban Patel[1], Mohammed Sufiyan[2], Nama Megana[3], Choppala Rohit[4]
[1,2,3,4]Computer Science and Engineering, CMR Technical Campus, Hyderabad, India)

How to cite this paper:
Punyaban Patel[1], Mohammed Sufiyan[2], Nama Megana[3], Choppala Rohit[4], "Text Summarization Using Word Frequencies", IJIRE-V3I03-396-368

Abstract: *Digital knowledge has become a crucial side of machine learning and is gift in large volumes on the web. To use this knowledge with efficiency, knowledge handling and process techniques are needed to filter data from documents and store them. Associate degree application of linguistic communication process, that helps in handling volumes of knowledge, is text summarization. Text summarization helps in condensation documents, and extract the necessary facts described in it. There are 2 techniques in text summarization: theoretical and extractive summarization. Extractive summarization extracts keywords from the document and combines them to supply a semantically incorrect outline, whereas, theoretical summarization produces a semantically correct outline of the text. During this paper, we have a tendency to compare completely different techniques to spot low and radio frequency words. We have a tendency to measure the techniques supported the right identification of positive and negative words.*

Key Word: *Text summarization; Abstractive summarization; Extractive Summarization; Semantic; NLP; Word Tokenization; Sentence Ranking*

I. INTRODUCTION

This project is termed as "Text Summarization Using Word Frequencies". This provides facility to cut back the massive volume of content by reducing it into a certain outline. This project uses language process to extract the outline from the given input text. We have a tendency to use a IP toolkit and regular expressions. The objective of this project is to grasp the ideas of language process and making a tool for a text account. The priority in Associate in Nursing automatic account is increasing therefore the manual work is removed. The project concentrates on making a tool which may mechanically summarize the document. With the growing quantity of information within the world, the interest within the field of automatic account generation has been wide increasing thus on reducing the manual effort of someone engaged on it.

II. MATERIAL AND METHODS

Reading the big content of text out there on-line is difficult for the users because it consumes an oversized quantity of your time. The projected system, Automatic Text report is way a lot of sensible and applicable in real time. The input text is pro cessed mistreatment linguistic communication process and processed input is born-again into vector kind mistreatment word embedding. Word embedding is that the collective name for a group of language modelling and have learning techniques in informatics wherever words or phrases from the vocabulary area unit mapped to vectors of real numbers. Sentence ranking is completed between sentences to extract higher hierarchic sentence that forms the extractive outline of the input.

Text Summarization: Text summary is the process of cutting down large publications into manageable paragraphs or sentences. The method collects critical information while maintaining the text's meaning. This cuts down on the time it takes to interpret vast amounts of data while avoiding omitting vital information, such as research articles. The process of constructing a concise, cohesive, and fluent summary of a text is known as text summarization. Highlighting the text's essential points is part of a longer text piece. Text summarization presents a number of issues, including text detection, text summarization, and text summarization. Interpretation and creation of a summary, as well as a review of the final summary in extraction-based summarizing, recognizing and leveraging key terms in the document. It's up to them to find helpful material to include in the summary.

Natural Language Processing: Text summarization is a very helpful and important component of Natural Language Processing (NLP). Let's begin with an explanation of what text summarizing is assume we have an abundance of text data in any form, including papers, magazines, and social media posts. We only need a summary of the material due to a shortage of time. We can summarize our text in a few lines by eliminating extraneous text and translating it into smaller semantic text form. In this method, we construct algorithms or programs that minimize the amount of the text and generate a summary of our text data. In machine learning, this is known as automatic text summarization. The practice of reducing the length of a document without losing its semantic structure is known as Text summarization.

Algorithm:

Input: A text in .txt or .rtf format.

Output: A relevant summarized text which is shorter than the original text keeping the theme or concept constant.

1. Read a text in .txt or .rtf format and split it into individual tokens.
2. Remove the stop words to filter the text.
3. Assign a weight value to each individual terms. The weight is calculated as:

$$WT = \frac{Frequency\ of\ the\ term}{Total\ no\ of\ terms\ in\ the\ document}$$

Procedure Methodology:

Text report approach relies on the removal of redundant sentences. A text report approach exploitation language process and varied extractive outline approaches like applied mathematics based mostly, topic-based, graph - based and machine learning based mostly. The options with higher results of extractive report is combined along to form higher report of the text.

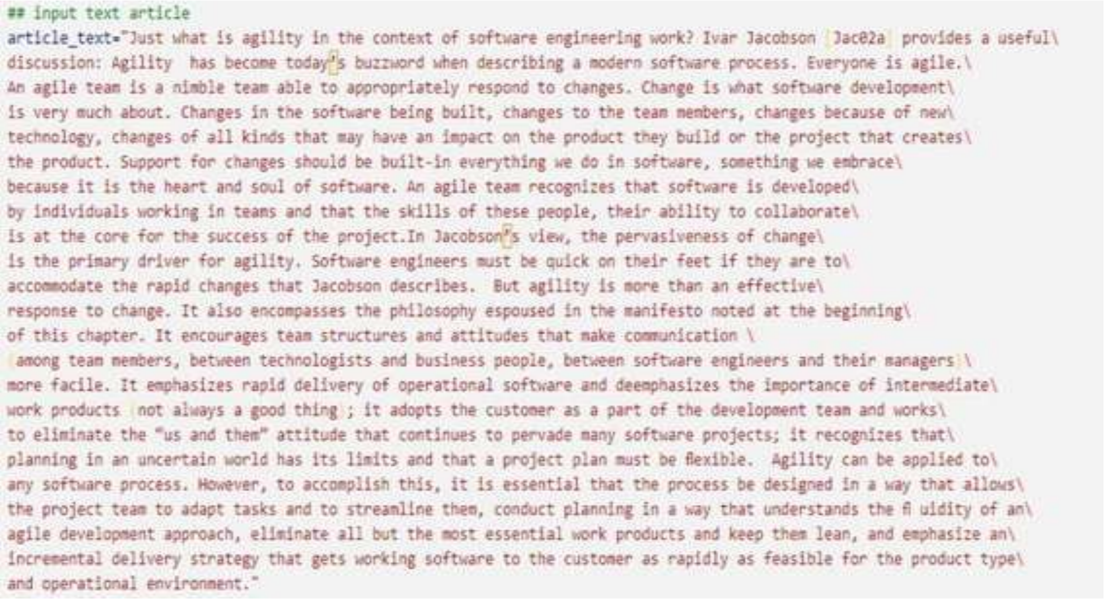
Text Summarization Steps:

1. Make sentences out of paragraphs.
2. Text Preprocessing.
3. Sentence Tokenization.
4. Calculate the Weighted Frequency of Occurrence.
5. In original sentences, replace words with weighted frequencies.
6. Arrange the sentences in descending order of their sum.

III. RESULT

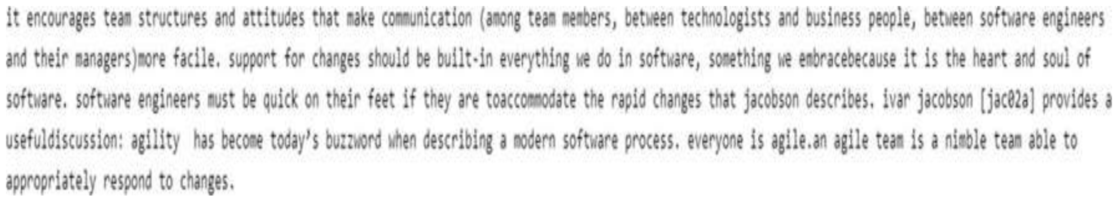
We've submitted the input file and performed data preprocessing, such as removing duplicates. We can observe the outcomes before and after applying the Natural Language Tokenization to sentences and words. Method for language processing wedelete extraneous sentences using NLP techniques. We first count the word frequencies, then rank the sentences based on the maximum word frequencies.

Figure no 1: Input



```
## input text article
article_text="Just what is agility in the context of software engineering work? Ivar Jacobson  Jac02a  provides a useful\
discussion: Agility  has become today's buzzword when describing a modern software process. Everyone is agile.\
An agile team is a nimble team able to appropriately respond to changes. Change is what software development\
is very much about. Changes in the software being built, changes to the team members, changes because of new\
technology, changes of all kinds that may have an impact on the product they build or the project that creates\
the product. Support for changes should be built-in everything we do in software, something we embrace\
because it is the heart and soul of software. An agile team recognizes that software is developed\
by individuals working in teams and that the skills of these people, their ability to collaborate\
is at the core for the success of the project.In Jacobson's view, the pervasiveness of change\
is the primary driver for agility. Software engineers must be quick on their feet if they are to\
accommodate the rapid changes that Jacobson describes.  But agility is more than an effective\
response to change. It also encompasses the philosophy espoused in the manifesto noted at the beginning\
of this chapter. It encourages team structures and attitudes that make communication \
 among team members, between technologists and business people, between software engineers and their managers \
more facile. It emphasizes rapid delivery of operational software and deemphasizes the importance of intermediate\
work products  not always a good thing ; it adopts the customer as a part of the development team and works\
to eliminate the "us and them" attitude that continues to pervade many software projects; it recognizes that\
planning in an uncertain world has its limits and that a project plan must be flexible.  Agility can be applied to\
any software process. However, to accomplish this, it is essential that the process be designed in a way that allows\
the project team to adapt tasks and to streamline them, conduct planning in a way that understands the fl uidity of an\
agile development approach, eliminate all but the most essential work products and keep them lean, and emphasize an\
incremental delivery strategy that gets working software to the customer as rapidly as feasible for the product type\
and operational environment."
```

Figure no 2: Output



```
it encourages team structures and attitudes that make communication (among team members, between technologists and business people, between software engineers
and their managers)more facile. support for changes should be built-in everything we do in software, something we embracebecause it is the heart and soul of
software. software engineers must be quick on their feet if they are toaccommodate the rapid changes that jacobson describes. ivar jacobson [jac02a] provides a
usefuldiscussion: agility  has become today’s buzzword when describing a modern software process. everyone is agile.an agile team is a nimble team able to
appropriately respond to changes.
```

IV. DISCUSSION

This paradigm can be applied to situations in which a large amount of data must be examined and a conclusion or output must be formed. Data is growing at an exponential rate nowadays. This massive amount of data creates a dilemma in terms of analyzing it and extracting various beneficial conclusions from it. This is where the concept of summarizing comes into play. Summarization is a technique for condensing a large dataset into little, countable lines of data so that a user may obtain something useful out of it in a short amount of time. Websites, blogs, news articles, webpages, and books can all benefit from this strategy. Aspreviously said, data is created from a variety of sources nowadays.

We can utilize this project to quickly comprehend an entire book and then construct our evaluation based on that summary.Researchers and scientists must read a large number of scientific publications and patents; a summary tool may help them save timeskimming through the articles, increasing productivity and assisting them in new discoveries. Lawyers have a vast number of case files to go through, and sifting through them all is a time-consuming task. This tool will aid in the summarization of these case file data, allowing the lawyer to grasp the case entirely in a short period of time.

V. CONCLUSION

The ideas and strategies utilized to construct trustworthy and understandable summaries using the cosine similarity sentence ranking algorithm are briefly described in this approach. This merits additional investigation in other market segments. The product'ssecurity will be determined by the outcomes of the evaluation. The primary goal of an autonomous summarization system is to provide an accurate and understandable summary from a huge amount of data. However, due to the vast amount of data available online in various formats, it still needs a lot of work.

Text summarization as a branch of NLP is quickly increasing due to the need for compressed and relevant synopses of a topic, as well as the massive volume of information available on the internet. Business analysts, government organizations, teachers, development researchers, marketing executives, and students can all benefit from text summarizing. Precise information improves the accuracy and efficiency of the search process. Users need this to process information in a short amount of time. This method can be employed in both the business and scientific worlds. It is less time consuming and less complicated than abstractive text summarization approaches, however the resulting summary is less accurate and meaningful than Abstractive methods.

References

[1] *Research Article An Automatic Multi document Text Summarization Approach Based on Naive Bayesian Classifier Using Timestamp Strategy.*

[2] *Research Article An Automatic Multi document Text Summarization Approach Based on Naive Bayesian Classifier Using Timestamp St.*

[3] *C. Lakshmi Devasena and M. Hemalatha, "Automatic Text categorization and summarization using rule reduction," IEEE-International Conference On Advances in Engineering, Science And Management (ICAESM -2012), Nagapattinam, Tamil Nadu, 2012, pp. 594-598.*

[4] *J. N. Madhuri and R. Ganesh Kumar, "Extractive Text Summarization Using Sentence Ranking," 2019 International Conference on Data Science and Communication (IconDSC), Bangalore, India, 2019, pp. 1-3.*

[5] *A. R. Pal and D. Saha, "An approach to automatic text summarization using WordNet," 2014 IEEE International Advance Computing Conference (IACC),Gurgaon, 2014, pp. 1169-1173.*

[6] *D. Hingu, D. Shah and S. S. Udmale, "Automatic text summarization of Wikipedia articles," 2015 International Conference on Communication, Information & Computing Technology (ICCICT), Mumbai, 2015, pp. 1-4.*

[7] *Steinberger, J., & Ježek, K. (2008). Automatic Text Summarization (The state of the art 2007 and new challenges). Znalosti, 30(2), 1-12.*

[8] *Yohei, S. (2002). Sentence extraction by TF/IDF and Position Weighting from newspaper articles. In Proceedings of the Third NTCIR Workshop.*

[9] *Das, D., & Martins, A. F. (2007). A survey on automatic text summarization. Literature Survey for the Language and Statistics, 3(3), 1-12.*

[10] *Santosh Kumar Bharti, “Automatic Keyword Extraction for Text Summarization in Multidocument eNewspapers Articles”, EuropeanJournal of Advances in Engineering and Technology, Vol. 4, Issue 6, pp. 410-427, 2017.*

[11] *Roshna Chettri, Udit Kr. Chakraborty, “Automatic Text Summarization”, International Journal of Computer Applications, Vol. 161, Issue 1, pp. 5-7, 2017.*