**TEXT SUMMARIZATION**

**ABSTARCT**

In multitudes of writers, poets and with their creative ideas that represent distinctively with different forms of writing style and forms in the literature that it becomes laborious to understand and analyze with challenging vocabulary, These can be solved by using AI.

Therefore, Our minor project aims to build an AI to influence in art field, focusing on poetry and creative writing. Our idea is to build an AI so that it can summarize the given text, picks up difficult words and sentences regarding the given text. Our minor project is helpful for several reasons to budding writers, poets and students of any degree. According to interviewed students of literature and avid readers, Their challenging part was to understand the strenuous text from various writers. Reports proved that India is one of the countries where people read the most. This idea would be the small steps in literature with giant leaps.

**CONTENTS**

CERTIFICATE

ACKNOWLEDGEMENT

ABSTRACT

CONTENTS

LIST OF FIGURES & TABLES

1. INTRODUCTION
   1. MOTIVATION
   2. OBJECTIVE
   3. SCOPE
2. LITERATURE SURVEY
3. ANALYSIS & DESIGN
4. IMPLEMENTATION

4.1 Real Time Operation

1. RESULTS

5.1Simulation Results

1. CONCLUSION AND FUTURE WORK
   1. Conclusion
   2. Future Work

**CHAPTER 1**

INTRODUCTION

1.1 MOTIVATION

The literature is expanding with the flow of poetry, prose, stories, novels and a wide range of forms increasing dramatically. Despite increasing its expansion, it has got its prominence in society. It would be a tiresome task for beginners and students. To subjugate this, we can summarize the text so that it can be beneficial. Therefore, we determined developing VIRTUAL EXPERT IN SUMMARISING LITERATURE, which can summarize the text with the help of Artificial Intelligence techniques.

We would like to implement a text summarization for literature, which is quite helpful to readers, learners and students. The most Significant areas of our field are Natural Processing Language(NLP), and Machine Learning develop our project-Virtual Expert In Summarising literature.

The subfield of summarization has been investigated by the NLP community for nearly the last half century. Radev et al. (2002) define a summary as “a text that is produced from one or more texts, that conveys important information in the original texts, and that is no longer than half of the original texts and usually significantly less than that”. This simple definition captures three important aspects that characterize research on automatic summarization:

• Summaries may be produced from a single document or multiple document

• Summaries should preserve important information,

• Summaries should be short.

Even if we agree unanimously on these points, it seems from the literature that any attempt to provide a more elaborate definition for the task would result in disagreement within the community. In fact, many approaches differ on the manner of their problem formulations

**1.2 OBJECTIVE**

According to the definition of summarization in introduction, Our objective built as “Summarization systems need to produce a concise and fluent summary, conveying the key information in the input.” In order to better understand, the operation of summarization systems as well as emphasis of summarization with the main techniques of Natural Language Processing, There several techniques that are distinguished: these favours summarisation with NLP

Syntactic analysis and semantic analysis are the main techniques used to complete Natural Language Processing tasks.

1. Syntax: refers to the arrangement of words and grammatical sentences.

In NLP, syntactic analysis is used to assess how the natural language aligns with the grammatical rules. Computer algorithms are used to apply grammatical rules to a group of words and derive meaning from them.

Here are some syntax techniques that can be used:

* [Lemmatization](https://en.wikipedia.org/wiki/Lemmatisation): It entails reducing the various inflected forms of a word into a single form for easy analysis.
* Word segmentation: It involves dividing a large piece of continuous text into distinct units.
* Part-of-speech tagging: It involves identifying the part of speech for every word.
* Parsing: It involves undertaking a grammatical analysis for the provided sentence.
* Sentence breaking: It involves placing sentence boundaries on a large piece of text.
* Stemming: It involves cutting the inflected words to their root form.

1. Semantics

Semantics refers to the meaning that is conveyed by a text. Semantic analysis is one of the difficult aspects of Natural Language Processing that has not been fully resolved yet.

It involves applying computer algorithms to understand the meaning and interpretation of words and how sentences are structured.

**Word Frequency**

The most prevalent discourse topic will play an important role in the summary and argue that lexical chains provide a better indication of discourse topic than does word frequency simply because different words may refer to the same topic. They define the strength of a lexical chain by its length, which is equal to the number of words found to be members of the same chain, and its homogeneity, where homogeneity captures the number of distinct lexical items in the chain divided by its length. They build the summary by extracting one sentence for each highly scored chain, choosing the first sentence in the document containing a representative word for the chain.

This strategy for summary selection—one sentence per important  
topic—is easy to implement.

**1.3 SCOPE**

The main principles of the VESL Project are using Natural processing language and Machine Learning. The scope of Virtual Expert In Summarising literature is to achieve maximum summarization of one particular genre in literature. This text summarization can be applied extensively in order to gain maximum benefit of long or indecipherable text. Summarising text with the help of NLP is increasing and flourishing in the coming years. It has become a reliable technology in the IT sector as well as interfering in many of the industries. It has become centre of gravity such as where Digital Information overloaded.

The advantage of this Linguistic method is: It is appealing because it offers perspectives for more semantically and linguistically rich treatment of text for summarization. Lexical chains help to capture all the sentences related to the central theme of the document providing the coverage of the topic, and thus cohesive summary is generated. The limitation of this method is: It does not consider domain-specific features and hence cannot be used for domain-specific summarization. It works efficiently for small documents.

Apart from summarising literature text, Summarization systems often have additional evidence they can utilize in order to specify the most important topics of the documents. For example, when summarizing blogs, discussions or comments are coming after the blog post that is excellent sources of information to determine which parts of the blog are critical and exciting. In scientific paper summarization, there is a considerable amount of information such as cited papers and conference information which can be leveraged to identify essential sentences in the original paper.

1. **LITERATURE SURVEY**

As part of the Literature survey, interaction with known personal working in machine learning domain and going through research papers is done and most optimum solution we came up with for the creation of the Text Summarization algorithm are using Frequency of tokens and using inbuilt sumy module in python.

The various research papers we have gone through gave us new routes into the stream of Natural Language Processing and lead us to uncover most fascinating facts about NLP and Machine learning and its use in communication and future technology.

1. **Analysis and Design :**

The required Text summarizer can be built using the concept of frequency Lexical ranking using Natural Language Tool kit and the other way is to use Sumy.py module.

***Code:***

*import sumy*

*from nltk.tokenize import punkt*

*import streamlit*

*import docx2txt*

*from sumy.parsers.plaintext import PlaintextParser*

*from sumy.nlp.tokenizers import Tokenizer*

*from sumy.summarizers.lsa import LsaSummarizer*

*from nltk.tokenize import sent\_tokenize*

*# taking input DATA*

*document1 = docx2txt.process("input\_text.docx")*

*# counting number of sentences present in original text*

*sent\_count = len(sent\_tokenize(document1))*

*#Using parsers PlaintextParser method:*

*# For Strings*

*parser = PlaintextParser.from\_string(document1,Tokenizer("english"))*

*# USING LSA*

*#Based on term frequency techniques with singular value decomposition to summarize texts.*

*#Let us summarize the text to 30% its original size or number of sentences*

*##Method using stopwords*

*from sumy.nlp.stemmers import Stemmer*

*from sumy.utils import get\_stop\_words*

*summarizer\_lsa2 = LsaSummarizer()*

*summarizer\_lsa2 = LsaSummarizer(Stemmer("english"))*

*summarizer\_lsa2.stop\_words = get\_stop\_words("english")*

*for sentence in summarizer\_lsa2(parser.document,sent\_count \* 0.30):*

*print(sentence)*

***Description:***

The above code is run after storing the text which is to be summarized in the “input\_text.docx” file and saving it.

After the code is run based on LSA method and term frequency the required output of summary is given.

The value 0.3 is considered here as t is found to be optimum for the textual data we onsidered to write summary. The size of the result can be increased, but this might lead to giving a summary which is similar to the size of the real text which defeats the purpose of summarization. Decreasing the value on the other hand will result in most frequenct terms in the result which might in some cases lead to improper summary.

.

**4. Implementation:**

***4.1. Real Time operation:***

This algorithm can be used to create a Summary for the given Text. Due to this it will help literature readers get the gist of a story before reading it in to the fullest. It will help students preparing for their theory exams to by giving them the most important points in the summary resulting in a quick revision. This will help the students to perform better in their preparations.

The real time operations and scope for text summarization project is vast. But as far as our project has progressed we can conclude with the above benefits from our implementation.

1. **RESULT:**

**5.1 Simulation Result:**

*Input:*

Let the following text be the input for the code then the resultant summary is as shown after the input.

Sasha was married to a fine young man about 5 years ago. After five years, Sasha was pregnant with her second child. They are a sweet happy family. Hearing this good news everyone was so happy. She and her husband knew it was a boy. Their first child was a girl. Pauline was four years old. She was a cute little princess who always tried to help her mother and be stated as a good kid. She always wanted to have a baby brother. She was excited about her new baby brother. She was going to help feed him. She was going to help dress him. She was going to help bathe him. She was very happy he was going to become an elder sister. "I'll teach him how to talk," she said. "I'll teach him how to walk. I'll teach him how to read my books." After a while it was time for Sasha’s delivery. Sasha went to the hospital. She gave birth to a healthy baby boy. Unfortunately, immediately after giving birth to the boy Sasha went through a terrible time. Then she got sick. It was an infection. The infection was growing inside her body it was eating her alive. Her body was attacking itself. There was no time for thinking, the doctors had to act fast. They amputated her legs. Then they amputated her arms. They saved her life. She was alive but lost both her arms and legs. Yet she was not sad. She was grateful. She had her husband. She had her daughter. She had her new son. "I'm alive!" she said. "I'm so happy just to be alive." Her first words after beating death made me realize; despite of what we go through one must always be happy with what they have. People sometimes always worry about what they want and stop realizing how much they have in their life they can be happy about. Sasha is a true inspiration to everyone of us

Output:

After the code is executed the output terminal is displayed with the following resultant summary.

Sasha was married to a fine young man about 5 years ago.

After five years, Sasha was pregnant with her second child.

Hearing this good news everyone was so happy.

She was a cute little princess who always tried to help her mother and be stated as a good kid.

After a while it was time for Sasha’s delivery.

She gave birth to a healthy baby boy.

Unfortunately, immediately after giving birth to the boy Sasha went through a terrible time.

The infection was growing inside her body it was eating her alive.

There was no time for thinking, the doctors had to act fast.

She was alive but lost both her arms and legs.

Her first words after beating death made me realize; despite of what we go through one must always be happy with what they have.

People sometimes always worry about what they want and stop realizing how much they have in their life they can be happy about.

1. **Conclusion and Future Work:**

**6.1 Conclusion:**

The concept of Natural Language Processing can be used to improve the Summarization technique and VESL project further into a better work. Giving out the summary for a given text readers to get a quick glance at what they are supposed to read, resulting in either letting them save time from reading the whole topic in required times or letting them understand the topic they are about to go through and let them know if that literature will peak their interest.

Result of our project displays a summary which has the important sentences which might be called to be in the bullet point format. This will let the readers who are going through the story after reading the whole text once prior to reading the summary go through a quick glance due to this the summary might be really beneficial to the person reading it.

**6.2 Future Work:**

This VESL project can be further extended where we can include the features where the genre of the text/story can be displayed. We can also include feature where all different words of specified parts of speech. This will help the users in improving their knowledge in grammer. Further mode the story can be made user interactable to search for a specific question or line in story to get the related information(this feature is already present in an existing application).