

NLP PROJECT

Text Summarizer

BY

GROUP NUMBER 07

PIDAKALA ABHINANDAN BABU

PREMA VIGNESH

SAAM PRASANTH DEEVEN PEDAPALLI

SRI LAKSHMI PRASANNA KONERU

Text Summarizer

Introduction

- Language is the primary communication tool used by everyone in daily life as a means to convey information.
- Predominantly information is stored in the form of text.
- In the earlier days, the data is stored and made available in printed and handwritten documents. Currently, these are all digitized.
- To retrieve the information, it is inevitable to depend on search engines to churn large volumes of documents to a specific user query which becomes a strenuous and time-consuming task.
- Search engines are using string matching algorithms with some statistical measures which short fall of semantic knowledge.

Cont....

- To overcome the hurdle, contextual summarization and intelligent search engines needed to be formatted.
- Searching the document with the document's context instead of word matching enhances document retrieval productivity.
- The information from the large documents is made available to these search engines at the semantic level in a succinct form.
- Text searching and summarization are two indispensable technologies that complement each other.

Motivation

NLP text summarizer is one of the cool applications of NLP (Natural Language Processing), which shortens the long document into a shorter one while retaining all important information from the document.

Nowadays, Short news is popular everywhere because people can get the important information in very short time, hence this saves time for the people. Abstractive summarization methods aim at producing summary by interpreting the text using advanced natural language techniques in order to generate a new shorter text.

Why Text Summarization ?

- Summarisation reduces reading time.
- When researching documents, summaries make the selection process easier.
- Automatic summarization improves the effectiveness of indexing.
- Automatic summarization algorithm are less biased than human summarizers.
- Personalized summaries are useful in question-answering systems as they provide personalized Information.
- Using automatic or semi automatic summarisation systems enable commercial abstract services to Increase the number of text documents they are able to process.
-

What is Text Summarisation ?

- Text summarization is the process of creating a short, coherent, and fluent summary of a longer text document and involves the outlining of the text's major points.
- Text identification, interpretation, summary generation, and analysis of the generated summary are some of the key challenges faced in the process of text summarization.
- The critical tasks in extraction-based summarization are identifying key phrases in the document and using them to discover relevant information to be included in the summary.

Categories of text summarisation

Extractive Summarization:

- These methods rely on extracting several parts, such as phrases and sentences, from a piece of text and stack them together to create a summary.
- Therefore, identifying the right sentences for summarization is of utmost importance in an extractive method.

Abstractive Summarization:

- These methods use advanced NLP techniques to generate an entirely new summary.
- Some parts of this summary may not even appear in the original text.

State of the art work on Text Summarizer

- 1) <https://buggyprogrammer.com/build-nlp-text-summarizer>

This Blog consists of “How To Build A Basic NLP Text Summarizer ”. In this,the basic Information and steps for Implementing Text summarization is available. Because of this Blog we got an Idea upon Text Summarizer.

- 2) <https://www.irjet.net/archives/V4/i8/IRJET-V4I8256.pdf>

Above two links are the references for our Project. We are Implementing our project (Text Summarizer) using NLP techniques at the Semantic level and Machine Learning concepts in development processes and NLP techniques like Lexical semantics and Selectional Restrictions and Word Sense Disambiguation.

How are we going to Implement this project ?

- The text summarizer implementation utilizes NLP techniques at the Semantic level and Machine Learning concepts in development processes.
- NLP techniques like Lexical semantics and Selectional Restrictions, Word Sense Disambiguation are used to understand and generate the text.
- We will also use Tokenization process.
- Clustering techniques (Density-based clustering) for "Formation of Sentence clusters concerning Tense Forms".
- Tense is used to retain original document tense in summarized text.

Work plan as per course plan

Project Task 1 (15%)

a. Presentation 1(Over Google Classroom):

Project Proposal/Literature Review (5%) (Aug Last week)

b. Presentation 2:

Midterm report (10%) (Sep 2nd Week/Oct 1st Week)

Project Task 2 (25%) (Nov 2nd/3rd week)

c. Project 2 implementation and presentation (17%)

d. Term paper(min 8 page double column) (8%)

Our work plan

For Presentation 1 on August 31st

Proposal for the project among group.

Collecting relevant information about the proposed project.

Presenting entire project idea as a first review. In this every one of our team is Involved.

We should learn new topics that are relevant for this project.

For Presentation 2

We will submit a midterm report for our project with relevant documents.

We will get a clear Information about project and number of ways to Implement this.

We will choose the best method for summarisation and Implement it as a trail run.

For Presentation 3 (Final Review)

We will submit the final project with proper documentation(Term Paper) and relevant slides and sources that we have used.



Thank you