Literature Review to Auto Summarization

1.1 Introduction

Text summarization is a method of summarizing data. A manual text summarization process is arguably an effective way to preserve the meaning of text. However, this is a time-consuming task. [1]

The process of condensing a portion of a text into a shorter version, minimizing the size of the original text while retaining important informational aspects and substantive meaning, is called an abstract.

Understand how the TextRank algorithm works and implement it in Python. Buckle up and this is going to be a fun ride. [3]

1.2 Types of Text Summarization

There is two main type of text summarization:

1.2.1 Abstractive summarization

Trying to understand the content of the text and then providing a summary based on that, which may or may not have the same sentences as present in the original test.

Abstractive summarization tries to create its own sentences and is definitely a step towards more human-like summaries.

The best method of abstractive is NLTK.

1.2.2 Extractive summarization

Extracted summaries, as their name suggests, contain formulations and phrases extracted from the original passage of text. There are various extraction techniques, but the most common and simplest is simply extracting sentences in the correct order from a passage of text.

Picking out sentences from the test that can best represent its summary.

It's more about learning to understand the importance of each sentence and their relations with each other rather than trying to understand the content of the text.

The best method of extractive is TextRank

1.3 Comparison between NLTK and TextRank

1.3.1 Advantage to NLTK

1. Wide range of tools and algorithms: NLTK comes with a number of tools and algorithms for tasks like tokenization, part-of-speech tagging, named entity recognition, and more.

2. Easy to use.

1.3.2 Disadvantage to NLTK

Performance: NLTK is implemented in Python, This can make NLTK somewhat slower than other toolkits that are implemented in faster languages such as C or C++.

2. Limited support for some languages: NLTK was developed primarily for working with English text, and while it does include support for other languages, this support may not be as extensive as for English.

3. Some features may be outdated.

4. May require more development time: Because NLTK is a general-purpose toolkit, you may need to spend more time developing.

1.3.3 Advantage to TextRank

1. It is simple to implement: TextRank is relatively easy to implement, as it does not require a lot of preprocessing or training data.

2. It is unsupervised: TextRank is an unsupervised learning algorithm, which means that it does not require any labeled data to work.

3. It can handle large datasets: TextRank is efficient and can handle large datasets, making it suitable for use on long documents or large collections of documents.

4. It can handle different languages: TextRank has been successfully applied to a variety of languages, including English, Arabic, Spanish, and Chinese.

5. It can extract keyphrases.

1.3.4 Disadvantage to TextRank

1. It relies on the quality of the input data: TextRank works by building a graph of the relationships between words or phrases in the text. If the input data is of poor quality, such as being poorly written or having a lot of noise.

2. It can be computationally expensive: Building the graph of relationships between words or phrases can be computationally intensive, especially for large datasets.

1.4 TextRank-Based Summarization

The modified TextRank proposed in this article uses the intuition behind the PageRank algorithm to rank sentences. This allows you to select the most important sentences from your input text document. In this model, we assigned an index to each sentence according to the input sequence of sentences in the document. The index value is heavily used to order the summarized sentences correctly to make the summary meaningful. In the next step, each sentence is tokenized into a series of words .

Represent each sentence in a given text document as a word vector to define the similarity between sentences. [6]

1.5 Similarity function

The number of common words between two sentences divided by the length of each to avoid promoting long sentences

1.6 TextRank model

At first, we separated the paragraph into sentences and put each one of them in a shape of a small circle.

Then, we connect between the similar sentence that is related to each other and removes the sentences that are not connected.

Next, we get the value of each sentence according to the number of edges outside it, each edge taking a value (number of sentences/1).

The sentences are displayed according to the number of sentences desired.

1.7 Conclusion

•TextRank models any document as an undirect graph using sentences as

nodes.

•A function to compute the similarity of sentences is needed to build edges in

between.

•This function is used to weigh the graph edges, the higher the similarity

between sentences the more important the edge between them.

•We can say that we are more likely to go from one sentence to another if they

are very similar.

2 Literature Review to course booking

E-learning has gotten better and better over time, and has brought about major changes in how we learn and who can achieve that learning. Education is growing rapidly, spurring the adoption of e-learning, which is a direct result of the integration of education and technology and is recognized as a powerful medium of learning. E-Learning is widely accepted in all educational sectors and academic institutions. In recent years, the application of e learning for teaching and learning has increased significantly, especially after lockdowns. eLearning has solved this problem because everyone has the knowledge and teaches them what they want, regardless of time difference or location.

Hot Course is an e-learning platform that provides educational content in various categories and levels. We will provide a new form of self-paced education that will change the way

education is for working adults and those who work short hours. With automatic summaries, Hot-Course cuts course length by more than half.

2.1 Advantages:

• Courses are summarized

• courses are offered in both English and Arabic

• Courses can be bought individually

• Easy to enroll as an educational center

• Past courses' reviews and rating

• Posts on courses with latest updates and discussion forums

2.2 Limitations:

• There is a possibility that the course you searched for is not available

4 SRS document

1 Introduction

1.1 Purpose

The program allows easy access to distinguished educational companies and the best courses at the best prices and discounts. Also, the user can summarize the large articles to make them smaller and easier to read to save his time and give him the same informational value.

1.2 Intended Audience and Reading Suggestions

We target students who need courses and companies that need these students.

1.3 Intended Use

For easy access to good educational courses without cost or effort, and I can summarize large books and articles.

1.4 Project Scope

The purpose of the program is to help students get access to the best educational companies, and we have thousands of educational companies in all fields and all over the world .

1.5 Definitions

Educational companies can join our world and upload their own courses and publications, and students can access excellent educational companies quickly without effort or cost, and they can see the opinions of others about educational companies

2 System Description

2.1 Methodology

The modified TextRank proposed in this article uses the intuition behind the PageRank algorithm to rank sentences. This allows you to select the most important sentences from your input text document. In this model, we assigned an index to each sentence according to the input sequence of sentences in the document. The index value is heavily used to order the summarized sentences correctly to make the summary meaningful. In the next step, each sentence is tokenized into a series of words .

At first, combine the articles into a single text, then do the preparation and processing of text to get rid of sentences that do not contain meaning.

All sentences are represented in a specific way, then chose the important sentence and represent it.

2.2 SYSTEM ANALYSIS & DESIGN

2.2.1 use case diagram

2.2.2 Sequence Diagram

2.2.3 ERD Diagram

2.3 Assumptions and Dependencies

2.3.1 Assumptions:

• there is a company's have access to login to the system and make an account to add courses and posts is already added by them

• the (id) for the company and student are auto generated by the system

2.3.2 Dependencies:

• To login the user needs to create an account

• To view courses available the user has to log in.

• To purchase a course a user has to choose the desired course and make a payment.

2.4 Intended Technologies

Platform: flutter, Django

Tools: APIs, library summarizer from module summa

Languages: dart, python

3 System Requirements

3.1 Functional Requirements

• Users must be able to create their own accounts.

• Users must be able to book courses.

• Users should be able to download files from courses or auto-summaries.

3.2 External Interface Requirements

3.2.1 User Interfaces

We have three main users:

3.2.1.1 Student

• login page

• create an account

• major's page

• home page

• student courses

• book page

• help page

• student profile

3.2.1.2 educational center

• login page

• create an account

• add a course

• educational center courses

• help page

• educational center profile

• home page

3.2.1.3 admin

• statistic page

• setting

• accept page

• home page

3.2.2 Hardware Interfaces

• Android

• iOS

• Windows

3.2.4 Communications Interfaces

This project supports all types of web browsers and android. We are using online payment.

3.3 System features

3.3.1 Auto summarization

3.3.1.1 Description and Priority

The auto-summarization feature provides a summary of the text that you provide.

3.3.1.2 Stimulus/Response Sequences

• He can see the old summaries he has summarized

• You can clear an old summary from history

3.3.1.3 Functional Requirements

• Users should be able to summarize large books and articles.

• Users should be able to download files after automatic summaries

3.3.1.4 Other system features include:

Download: Download the book or text after summarizing

New summarize: to do new summarize as free cost

3.3.2 signup

3.3.2.1 Description and Priority

Providing a feature for the user to create a new account to save his data and refer to it at any time.

3.3.2.2 Stimulus/Response Sequences

•Can create an account with google or Facebook

•You can delete the account if you want

3.3.2.3 Functional Requirements

•Users should be able to create any number of private accounts.

•Users should be able to log in from anywhere on any type of device.

3.3.2.4 Other system features include:

major: choose your favorite fields

locations: choose all locations

3.4 Nonfunctional requirements

Performance

✓ Ensure that System features are highly responsive in a very short time so that the system works under high efficiency with an average time (3 seconds)

✓the system and database can handle pdf files

✓database able to meet the requirements of continuous data growth continuously to ensure fast response of data presentation

Safety and secure

✓ If there is extensive damage to a wide portion of the database due to catastrophic failure, such as a disk crash, the recovery method restores a past copy of the database that was backed up to archival storage

✓user cannot use previously used password as new one

✓the system never save the payment methods information such as credit or debit cards

✓the system will always backup itself for easy restoration and data protection

Quality ✓ the system should auto -summarize the i nput in an understandable way that includes all important points that were originally mentioned in the original text.