Contents

Abstract (300 words)

In the immense data era, there has been an explosion within the quantity of textual information from a spread of sources. This volume of textual content is an incomputable offer of statistics and know-how that needs to be with success summarized to be useful. This growing handiness of files has demanded complete analysis within the natural language processing space for automatic text summarisation.Text summarization is the procedure of automatically growing a shorter version of 1 or greater text documents. It is a crucial way of locating relevant statistics in massive textual content libraries or in the Internet.

Essentially, text summarization techniques are categorized as Extractive and Abstractive. Extractive techniques perform text summarization by using selective sentences of files consistent with some criteria. Abstractive summaries try to improve the coherence amongst sentences by removing redundancies and clarifying the context of sentences.

In phrases of extractive summarization, sentence scoring is the method maximum used for extractive textual content summarization. This project describes and performs a quantitative and qualitative evaluation for sentence scoring available in the literature. Three different datasets (News, Blogs and Article contexts) have been evaluated and summarized to the number of lines desired by the user.. In addition, we have also provided a video to text summarizer which is able to summarize informational Youtube videos into a meaningful summarized paragraph by taking the Youtube video ID as an input from the user.

Introduction

Automatic summarisation means that a automatically short output is dependent once a input is applied. We should remember that an input could be a well structured document. For this even there are opening pre-processes like Tokenization, Sentence Segmentation, Removing stop words and Word Stemming. An extractive summarisation method consists of selecting most vital sentences, words, paragraphs etc. from the first record and concatenating them into shorter kind. An Abstractive summarization is a grasp about the fundamental ideas in a report and expresses those thoughts into an apparent simplistic language. Text Summarization is a lively difficulty of research amongst every text regarding the IR and NLP groups.People can hold up with the arena affairs via listening to information bites. People can visit the films in large part over the idea

of evaluations they've visible. People can base funding decisions on stock market updates. With summaries, People can make powerful decisions in less time. The motivation right here is in

conformity to assemble a sure system which is computationally environment friendly and creates easily understandable and meaningful summaries.

The Project is based on **the Text Rank Algorithm** to summarize the articles whose urls are provided by the end-user.

Text Rank is a textual content summarization method which is utilized in Natural Language Processing to generate Document Summaries.

It uses an **extractive approach** and is **an unsupervised graph** based textual content summarization method. It is based on the Page-Rank technique used by search engines like google to provide the top prioritized pages or links to the end -user based on his/her search using this ranking algorithm to calculate the rank of the web-pages.
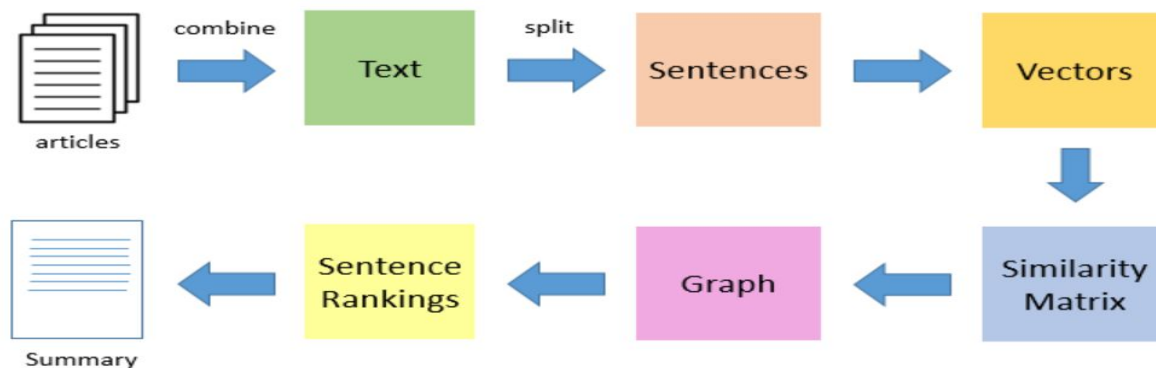
The Text Rank Algorithm works in the following ways:

It extracts all of the sentences from the text document, either with the aid of splitting at the whitespaces or full stops (and other stopwords), or any other way in which you desire to define your sentences.
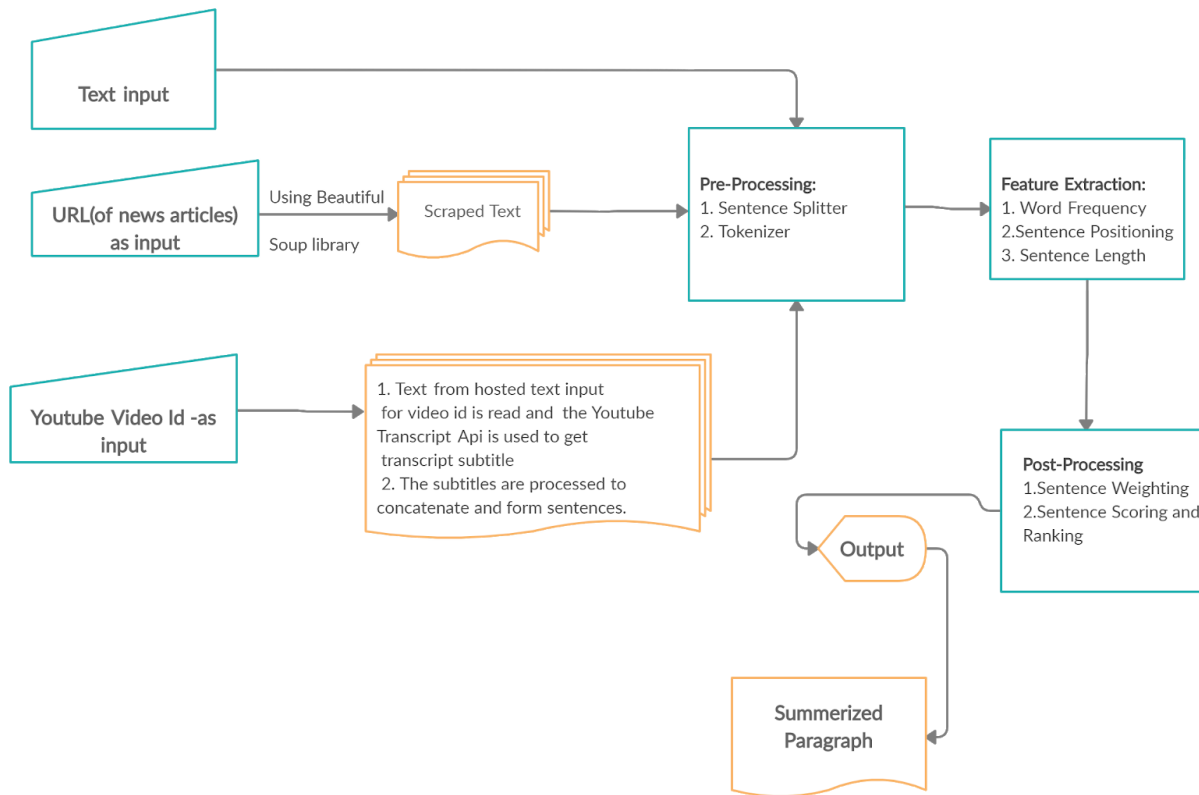
A Graph is then generated out of the sentences.The nodes represent the sentences, whilst the load on the edges among two nodes is located via the use of a Similarity function, like Cosine Similarity or Jaccard Similarity.

The next step involves finding significance (scores) or rank of every node through iterating the algorithm till convergence, (i.e) till constant ratings are obtained. It can be said that this step includes utility of PageRank algorithm to the document, with the only difference being that the nodes are sentences in preference to web pages.

Finally, the sentences are looked after in a descending order primarily based upon their scores. The first ok sentences are selected to be a part of the textual content summary.

Architecture diagram

Text input

URL(of news articles) as input → Using Beautiful Soup library → Scraped Text

Youtube Video Id -as input → 1. Text from hosted text input for video id is read and the Youtube Transcript Api is used to get transcript subtitle 2. The subtitles are processed to concatenate and form sentences.

Pre-Processing:
1. Sentence Splitter
2. Tokenizer

Feature Extraction:
1. Word Frequency
2. Sentence Positioning
3. Sentence Length

Post-Processing
1. Sentence Weighting
2. Sentence Scoring and Ranking

Output

Summerized Paragraph

Background study (Related papers and study)

In paper, *Text summarization techniques: a brief survey by Allahyari,* main approaches to automatic text summarization are described. They reviewed varied processes for summarization and described the advantages and disadvantages of different methods.

We also studied *A Context Based Text Summarization System by Ferreira,* This paper tries to find the best combinations of sentence scoring methods for three kinds of documents: news, blogs, and articles.

Leveraging BERT for Extractive Text Summarization on Lectures by Derek Miller in his paper reports on the project called "lecture summarization service", a python-based RESTfulservice that utilizes the (Bidirectional Encoder Representations from Transformers) BERT model for text embeddings and K-Means clustering to identify sentences closest to the centroid for summary selection. The purpose of the service was to provide student's a tool to summarize lecture content, based on their desired number of sentences. While the results of this paper by Derek utilizing BERT for extractive text summarization were promising, there were still areas where the model struggled, providing future research opportunities for further improvement.

The objective of *Summarizing News Paper Articles: Experiments with Ontology-Based, Customized, Extractive Text Summary and Word Scoring by Kallimani* paper is to save

prospective readers' time and effort in finding the useful information in a given huge article. Various analyses of results were also discussed by comparing it with the English language. Only extractive summarization method is considered and developed for the study.

A Survey on Extractive Text Summarization by Moratanch talks about the various techniques, populous benchmarking datasets and challenges of extractive summarization have been reviewed. This paper interprets extractive text summarization methods with a less redundant summary, highly adhesive, coherent and depth information. In this work, a comprehensive review of extractive text summarization process methods has been ascertained.

By studying these papers,we tried to develop a 1 stop summarizer which can summarize news article, blog, simple text and even a youtube video. Video summarizer study was not as frequent as other papers. So we tried to incorporate this new feature in our project.

Methodology (Explanation about algorithm, methods, datasets)

Hindu Article Summarizer Module

Libraries Used:

1) Beautiful Soup

2) sklearn

Explanation/Procedure:

1)Given a Hindu News Article link and the number of sentences to be present in the summary as input, using BeautifulSoup library, the website gets scraped and it returns string back to the variable.

2)Tokenize the entire string into sentences and sentences into words. We need unique words in order to determine their comparative frequency in the document, and give them a relative score and rank; we need individual sentences to subsequently sum the scores of each word within in order to determine 'sentence score and rank''.

3)After counting the occurence of each token in the string. After finding the counts, we build a frequency distribution of words.

4)Further ahead, give a rank to the sentences by using the frequency table processed. This is simply adding all the relative scores of each token in a sentence. This func takes a max_len arg which sets a max length to groups of tokens which are to be added for use in the summarization.

5)Based on the scores and ranks, select the top 'k' sentences that represent the gist of the article. Display the summarized version along with the top 'k' sentences and their sentence scores.

WikiPedia Summarizer

Similarly in WikiPedia Summarizer, Beautiful Soup library is used to scrape the article. Then it follows the same algo as followed in Hindu Article Summarizer.

Libraries used:

1) Gensim Summarizer

2) Beautiful Soup

3) nltk.corpus

4) nltk.tokenise

Explanation/ Procedure:

1) Obtain Text from hosted text input.

2) Extract kewords

3) Create Summary

4) Print the summary

Text Summarizer

Libraries Used:

1) nltk.tokenize

2) nltk.corpus

3) nltk.probability

4) heap

Explanation/Procedure:

1) Read the text from hosted text input

2) Sanitize input from tabs, spaces unwanted new lines

3) Tokenise content

4) Score tokens

5) Summarize from ranks and tokens.

Youtube Video Summarizer

Libraries Used:

1) nltk.tokenize

2) nltk.corpus

3) nltk.probability

4) heap

5) youtube_transcript_api

Explanation/Procedure:

1) Read the text from hosted text input for video id

2) Use the Youtube Transcript Api to get transcript subtitle

3) Process the subtitles to concatenate and form sentences.

4) Sanitise text from tabs, spaces unwanted new lines

5) Tokenise content

6) Score tokens

7) Summarise from ranks and tokens.

Proposed model (Explanation with diagram)

In this Project we thus propose a single platform to find the summaries of different news articles (here, included wikipedia articles , hindu articles) and also long texts.We have also added the Video Summarizer feature to this project which is not so commonly provided by other platforms.

The main idea behind this model is to reduce the time consumed by people to read the articles and understand , we propose this platform to quickly get the gist of the article and in less time understand the important content present in the article. Extractive Text Summarization has been used to generate the summaries of the articles based on the Text Rank Algorithm- the entire string is tokenized into sentences and sentences into words. We need unique words in order to determine their comparative frequency in the document, and give them a relative score and rank; we need individual sentences to subsequently sum the scores of each word within in order to determine 'sentence score and rank''. Finally , based on the rank the top 'k' sentences are given as the summarized output.

This is quite helpful for summarizing long news articles , informational videos of long duration and huge texts to understandable and meaningful summary to increase the convenience and ease of the end-users in gathering information and useful data in much less time.
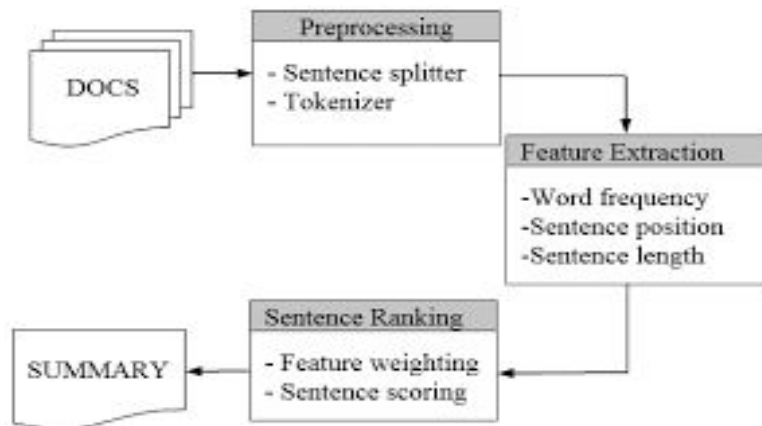

Fig: Proposed Model

Results and Discussion

Sample Tokenise Content Code

```python
def tokenize_content(content):

    stop_words = set(stopwords.words('english') +
list(punctuation))

    words = word_tokenize(content.lower())

    return (sent_tokenize(content), [word for word in words if
word not in stop_words])
```

Sample Assigning Tokens a Score Code Snippet

```python
def score_tokens(sent_tokens, word_tokens):

    word_freq = FreqDist(word_tokens)

    rank = defaultdict(int)

    for i, sent in enumerate(sent_tokens):

        for  word in word_tokenize(sent.lower()):
```

```python
        if word in word_freq:

            rank[i] += word_freq[word]

    return rank
```

Sample Summarize Code Snippet

```python
def summarize2(ranks, sentences, length):


    if int(length) > len(sentences):

        print('You requested more sentences in the summary than
there are in the text.')

        return ''

    else:

        indices = nlargest(int(length), ranks, key=ranks.get)

        final_summary = [sentences[j] for j in indices]

        return ' '.join(final_summary)
```

Testcase Number	Input	Output
1	https://www.thehindu.com/news/national/pm-modi-congratulates-new-zealand-counterpart-jacinda-ardern-on-poll-win/article32885939.ece	Also Read Jacinda Ardern wins landslide re-election in New Zealand vote Jacinda Ardern wins landslide re-election in New Zealand vote In his tweet, Mr. Prime Minister Narendra Modi on Sunday congratulated his New Zealand counterpart Jacinda Ardern for securing a second term in office, and said he looked forward to working together for taking the bilateral relationship to a higher level.
2	https://en.wikipedia.org/wiki/Python_(programming_language)	Python uses dynamic typing and a combination of reference counting and a cycle-detecting garbage collector for memory management. It also features dynamic name resolution , which binds method and variable names during program execution. Python's developers strive to avoid premature optimization, and reject patches to non-critical parts of the CPython reference

		implementation that would offer marginal increases in speed at the cost of clarity. When speed is important, a Python programmer can move time-critical functions to extension modules written in languages such as C, or use PyPy, a just-in-time compiler.
3	https://www.youtube.com/watch?v=**KR0g-1hnQPA** here we take video youtube id	t's the new LiDAR scanner on iPhone 12 Pro.Now back to Joz.As you can see, iPhone 12 Pro is loaded w. ith incredible innovations.And when you combine these advanced capabilities with 5G,it will enable y. ou to do so much more on the go.And 5G is going to unlock even more opportunities for our developers. Fusion to all four camerasof iPhone 12 Pro,including the front-facing TrueDepth camera.We combine st. ate-of-the-art camera hardwarewith this amazing compute powerto create a pro camera system that is u. nlike any other device.iPhone 12 Pro features our amazing Ultra Wide camerawith a 120-degree field o. f view,and our new Wide camera with faster f/1.6 apertureand 7-element lens that lets in 27% more li.
4	Sample Large text	Specifically, the wee things that we see as part of graphics, maps, visualizations (wee things in space) as well as the wee things we experience as part of interactions, navigation, and usability (wee things in time). This means everything from sequences of small graphics that help us make comparisons, to tiny locator maps that help orient us within a larger graphic, to navigation icons that give hints about how we should make our way around a page. So with Waldo as our anti-hero, let's take a look at how people read and interpret small visual forms, why tiny details can be hugely useful, and what principles we can apply to make all these little images and moments work for us as designers. Wee Things In Space Probably the most immediate definition of wee things are things that are physically small: little things on a page. But while he looms large in our imagination, our childhood searches for Waldo typically stayed pretty small – Waldo is a tiny person in the middle of lots of other tiny things.

1Stop Shortener

Get summaries for your papers, articles, homework etc.

Made By Piyush, Aditi and Chaitty

Original text:

Prime Minister Narendra Modi on Tuesday said his government's 'historic' farm reforms will create opportunities to take farmers to entrepreneurship and added his regime is focused on increasing their income. Also read: Mann ki Baat | Farmers can break free of APMC shackles: PM Modi Mr. Modi was speaking after releasing the autobiography of former Union Minister Balasaheb Vikhe Patil and renaming the Pravara Rural Education Society in Ahmednagar district after the Padma Bhushan recipient, via video conferencing. Terming the farm reforms as historic, the Prime Minister said, "Today opportunities are being created to take farming and farmers from the role of annadata (food provider) to entrepreneurship." Referring to high milk, sugar, and wheat production in Gujarat, Maharashtra, Haryana and Punjab, Mr. Modi said such models of local enterprise will take the country forward. Also read: Comment | Farm Acts – unwanted constitutional adventurism He said there was a time after Independence when the country did not have sufficient food stock. "The governments priority then was to increase food production. So the entire focus was on increasing production. "Farmers toiled hard to achieve the objective," Mr. Modi said. "But governments and policies did not pay attention towards profitability of farmers while being concerned about increasing productivity. "People forgot about income for farmers. But for the first time, this thinking has been changed," the Prime Minister said and listed steps the Centre has taken for the welfare of farmers. Noting that the COVID-19 threat still persists, Mr. Modi called for observing social distancing norms and maintaining effective personal hygiene.

Summarised text:

Prime Minister Narendra Modi on Tuesday said his government's 'historic' farm reforms will create opportunities to take farmers to entrepreneurship and added his regime is focused on increasing their income. Terming the farm reforms as historic, the Prime Minister said, "Today opportunities are being created to take farming and farmers from the role of annadata (food provider) to entrepreneurship. Modi was speaking after releasing the autobiography of former

Fig Above: Here Hindu Article gets summarized

Choose the no. of sentences in the summary

300   −   +

Summarize WikiPedia Article

The World Food Programme is the food-assistance branch of the United Nations and the world's largest humanitarian organization focused on hunger and food security. Founded in 1961, it is headquartered in Rome and has offices in 80 countries. As of 2019, it served 97 million people in 88 countries, the largest since 2012, with two-thirds of its activities conducted in conflict zones. In addition to emergency food relief, WFP also offers technical assistance and development aid, such as enhancing capacity for emergency preparedness and response, managing supply chains and logistics, promoting social safety nets, and building resilience against climate change. It is an executive member of the United Nations Development Group, which collectively aims to fulfil the 17 Sustainable Development Goals , with a priority on achieving SDG 2 for "zero hunger" by 2030. WFP launched its first programmes in 1963 by the FAO and the United Nations General Assembly on a three-year experimental basis, supporting the Nubian population at Wadi Halfa in Sudan. The Logistics Cluster is an Inter-Agency Standing Committee humanitarian coordination mechanism whose primary role is supporting emergency responses. One of eleven sectoral coordination bodies, it was established by UN General Assembly resolution 46/182 in December 1991 and extended in the Humanitarian Reform of 2005, with new elements adopted to improve capacity, predictability, accountability, leadership and partnership. The Logistics Cluster provides coordination and information management services to support operational decision-making and improve the predictability, timeliness and efficiency of humanitarian emergency responses. Due to its expertise in the field of humanitarian logistics, the World Food Programme was chosen by the IASC as the lead agency for the Logistics Cluster. WFP also acts as a "provider of last resort" offering common logistics services, when critical gaps hamper the humanitarian response.

1Stop Shortener

Get summaries for your papers, articles, homework etc.

Made By Piyush, Aditi and Chaitty

Choose the no. of sentences in the summary

10   −   +

Summarize Text

Specifically, the wee things that we see as part of graphics, maps, visualizations (wee things in space) as well as the wee things we experience as part of interactions, navigation, and usability (wee things in time). This means everything from sequences of small graphics that help us make comparisons, to tiny locator maps that help orient us within a larger graphic, to navigation icons that give hints about how we should make our way around a page. So with Waldo as our anti-hero, let's take a look at how people read and interpret small visual forms, why tiny details can be hugely useful, and what principles we can apply to make all these little images and moments work for us as designers. Wee Things In Space Probably the most immediate definition of wee things are things that are physically small: little things on a page. But while he looms large in our imagination, our childhood searches for Waldo typically stayed pretty small – Waldo is a tiny person in the middle of lots of other tiny things. Here are a bunch of examples of small multiples in the wild – maps and planets, first lady hair styles and telegraph signals, food trucks, fashion color trends and dressing appropriately for different climates, the distribution of deaths in the 1870's and last but not least, Bill Murray's hats. They do this by recording information for us to make use of later, lending a hand to our (pretty terrible) working memories, helping us search and discover and recognize. This "spotlight" of attention might be good at letting us concentrate, but it makes us pretty bad at spotting Waldo. Instead, he forces us to use our attentive processing: the conscious, slow, sequential process that lets us focus on one thing at a time. In many ways, Waldo is a great example of what NOT to do when using wee things in your own work.

1Stop Shortener

Get summaries for your papers, articles, homework etc.

Made By Piyush, Aditi and Chaitty

Writing Youtube Video Id in above figure. And in figure below, we get result for youtube video summarized

Conclusion

In this project we have created a text summarizer for summarizing Hindu Articles, WikiPedia Articles, text paragraphs and more importantly Youtube Video Shortener. We believe hosting this project online will help a lot of people. Youtube Video Summarizer can help students understand main points from a lecture, match etc. Three different datasets (News, Blogs and Article contexts) have been evaluated and summarized to the number of lines desired by the user.. In addition, we have also provided a video to text summarizer which is able to summarize informational Youtube videos into a meaningful summarized paragraph by taking the Youtube video ID as an input from the user.

References (Harvard style, sort by name)

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez, J.B. and Kochut, K., 2017. Text summarization techniques: a brief survey. *arXiv preprint arXiv:1707.02268*.

Ferreira, R., de Souza Cabral, L., Lins, R.D., e Silva, G.P., Freitas, F., Cavalcanti, G.D., Lima, R., Simske, S.J. and Favaro, L., 2013. Assessing sentence scoring techniques for extractive text summarization. *Expert systems with applications*, *40*(14), pp.5755-5764.

Kallimani, J.S., Srinivasa, K.G. and Reddy, B.E., 2012. Summarizing news paper articles: experiments with ontology-based, customized, extractive text summary and word scoring. *Cybernetics and Information Technologies*, *12*(2), pp.34-50.

Miller, D., 2019. Leveraging BERT for extractive text summarization on lectures. *arXiv preprint arXiv:1906.04165*.

Moratanch, N. and Chitrakala, S., 2017, January. A survey on extractive text summarization. In *2017 international conference on computer, communication and signal processing (ICCCSP)* (pp. 1-6). IEEE.