



Ain Shams University
Faculty of Computer & Information Sciences
Computer Science Department

Arabic Text Summarization

By:

Engy Assem Ahmed
Aya Ashraf Mostafa Selim
Omnia Ayman Mohamed
Lujain Ahmed Elsayed
Ahmed Mohamed Elsayed Soliman

Computer Science Department,

Under Supervision of:

[Dr.Mohamed Maborok]
Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

[TA.Youmna Ahmed]
Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

July 2022



Ain Shams University
Faculty of Computer & Information Sciences
Computer Science Department

Arabic Text Summarization

By:

Engy Assem Ahmed [CS]
Aya Ashraf Mostafa Selim [CS]
Omnia Ayman Mohamed [CS]
Lujain Ahmed Elsayed [CS]
Ahmed Mohamed Elsayed Soliman [CS]

Under Supervision of:

[Dr.Mohamed Maborok]
Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

[TA.Youmna Ahmed]
Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

Abstract

Text summarization is the process of producing a shorter version of a specific text. Automatic summarization techniques have been applied to various domains such as medical, political, news, and legal domains proving that adapting domain-relevant features could improve the summarization performance. Despite the existence of plenty of research work in the domain-based summarization in English and other languages, there is a lack of such work in Arabic due to the shortage of existing knowledge bases.

The project approach is to summarize Arabic text. Arabic text summarization is one of the natural language processing applications which aims to reduce the original text and return the important information from the original text. Which aims to reduce the time that the reader takes to read the article and provide him with the same important information in the original text.

Arabic language has a complex morphological structure which makes it difficult to extract nouns, so we use a morphological analyzer that is used to solve the problem of nouns extraction and use nouns extraction to build a graph as initial rank, and cosine similarity was used to weigh the edges between sentences. In summary extraction to prevent redundancy, if the overlapping between the selected sentence and any other sentence in the summary is very high then, this sentence is neglected. A Modified PageRank algorithm was used to extract the summary, this algorithm was used by making the initial rank of the sentence as the number of nouns it has, and the weight of the edge is the cosine similarity between the connected nodes. To evaluate the performance of this approach EASC Corpus is used as a standard.

Table of Contents

Abstract.....	i
List Of Algorithms.....	iii
List of Figures	iv
List of Abbreviations	v
1- Introduction	1
1.1 Motivation.....	1
1.2 Problem Definition.....	2
1.3 Objective	2
1.4 Time Plan.....	3
1.4 Document Organization	4
2- Background	5
3- Analysis and Design.....	8
3.1 System Overview.....	8
3.1.1 System Architecture.....	8
3.1.2 System Users	9
3.2 System Analysis & Design	10
3.2.1 Use Case Diagram	10
3.2.2 Class Diagram	11
3.2.3 Sequence Diagram	13
3.2.4 Database Diagram.....	13
No database was needed for this project.	13
4- Implementation and Testing.....	14
5- User Manual.....	23
6- Conclusion and Future Work	26
6.1 Conclusion.....	26
6.2 Future Work.....	27
References	28

List Of Algorithms

Preprocessing Algorithm1	14
Morphology Algorithm 1	15
Building Graph Algorithm 1	16
Modified PageRank Algorithm 1	17
Summary Extracriion Algorithm 1	18
Calculate Accuracy Algorithm 1	19
Extract Title Algorithm 1	21

List of Figures

Figure 1. Time Plan.....	3
Figure 2. System Architecture	8
Figure 3. Use Case diagram	10
Figure 4. Class diagram	11
Figure 5. Sequence diagram.....	13
Figure 6. Graph	16
Figure 7. Evaluation results	20

List of Abbreviations

<u>Abbreviation</u>	<u>What the abbreviation stands for</u>
CSS	Cascading Style Sheets
HTML	Hyper Text Markup Language
NER	Named Entity Recognition
NLP	Natural Language Processing
Rouge	Recall-Oriented Understudy for Gisting Evaluation
RST	Rhetorical Structure Theory
TF-IDF	Term Frequency-Inverse Document Frequency

1- Introduction

1.1 Motivation

Due to the massive amount of data written every single day in a different format on the internet since its invention two decades, the desire for automatic text summarization to extract the most important information from the document intensified.

As people need more time to read the whole text of many documents with the same topic to get the main idea, it is most important to provide an improved mechanism to extract the information quickly and most efficiently. text summarization came to overcome this issue, as it saves time and effort by generating a brief text edition of the source text containing the same ideas without needing to read the whole text. Also, it saves the cost compared with human expert summarization.

There are many pieces of Research in NLP for Latin languages compared to the Arabic language, especially in the summarization field. Many text summarization researches were conducted for the English language because it is a simple language in its structure and grammar, which is the opposite of the Arabic language, as it has a complex morphological structure which makes it very difficult to extract nouns to be used as a feature for summarization process. Arabic text summarization still suffers from the low performance and the lack of research done in this application of natural language processing.

There are more than 350 million people who speak Arabic around the world. It is very difficult for those people to manually extract the summary of large documents of text.so, Arabic text summarization is widely required because of the huge number of people that will benefit from an application like this.

1.2 Problem Definition

Arabic text summarization is one of the challenging open areas for research in natural language processing (NLP) and Researches on forming Arabic text summaries have not been done sufficiently when compared to the research accomplished in English or other languages, this is due to some issues and challenges that slow down the progress in Arabic Natural Language Processing.

The Arabic language has different problems, represented the Arabic language is a highly derivational and inflectional language, which makes morphological analysis such as lemmatization and stemming a very complex task, the absence of Diacritics that is integral in Arabic texts increases the complexity of inferring s' meaning, also a language lacks capitalization leading to a great challenge in the process of Named Entity Recognition (NER) system, so it is considered highly ambiguous in comparison to other languages.

The shape of letters in the Arabic language differs if the letter is in the beginning or the middle or the end of the word, also the Arabic language has three varieties: classical Arabic (found in religious scripts), modern standard Arabic (found in today's written Arabic and spoken in formal channels), and colloquial or dialectal Arabic (the spoken language), also Arabic dialects vary from one Arab country to another, All these challenges may affect the results of the next process of the text such as summarization, classification, for sentiment analysis these problems and challenges, it led to the difficulties of working in the Arabic language. the poverty of research in it let to the lack of automated Arabic NLP tools such as lexicons, semantic role labelers, and named entity recognition complicate the process more.

1.3 Objective

Create an application for Arabic text summarization which will:

- Generate short text.
- Present the most important content.
- Extract title for the text.

1.4 Time Plan

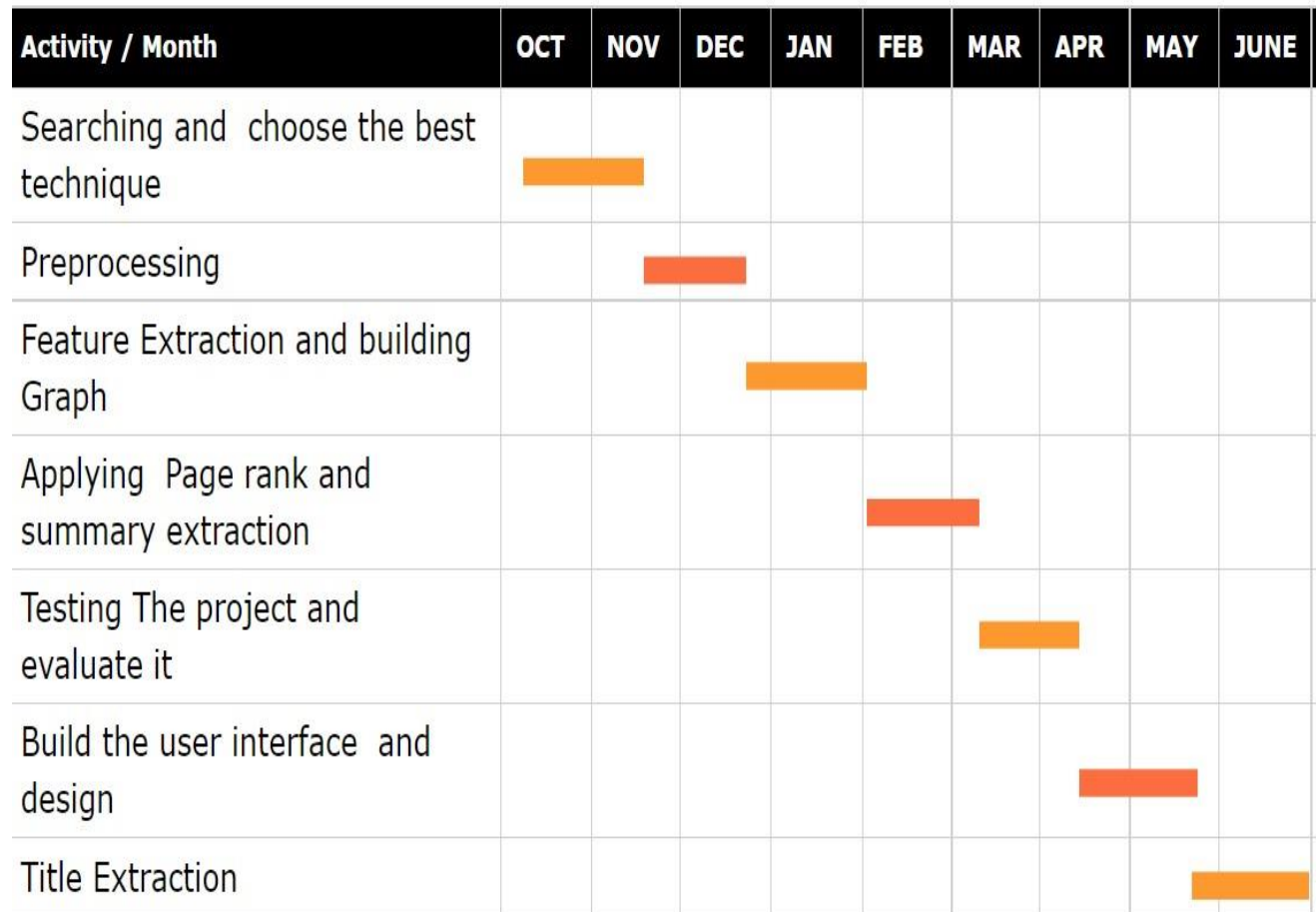


Figure 1. Time Plan

1.4 Document Organization

In chapter 2 The background was discussed, in this section we introduced: the field of the project and the difference between extractive and abstractive text summarization, the different techniques for text summarization and a Comprehensive Survey about Arabic Text Summarization techniques.

In chapter 3 We discussed System overview

- System architecture and description of system phases.
- Describe system users: who can use this application, and user's characteristics.

And System analysis and design:

- Use case diagram and description for use cases.
- Class diagram and classes description.
- Sequence diagram.

In chapter 4 we discussed the description of main functionalities, algorithms, and creation of pseudo-code for these functions.

Also the description of the dataset used has been discussed in addition to the description of Evaluation matrices and testing. Finally, UI design is discussed.

In chapter 5 we have the user manual that describes how a user can run the application.

In chapter 6 we have 2 sections one for Conclusion (Summing up document and project result) and the other for Future work (Describe features that can be added to the project).

2- Background

There are three types for text summarization extractive, abstractive and hybrid.

Extractive text summarization: generate a summary by selecting the most important sentences in the text by their score, without changing words in the original text, there are many approaches for extractive vary from simple and complex semantic processing.

Abstractive text summarization: generates a summary like a human's summary with new sentences by paraphrasing the original text.

Hybrid text summarization: it is a combination between Extractive and Abstractive.

Extractive text summarization techniques:

- Graph-Based approach:
This method is used to avoid redundancy in the text by using sentence-based graph to describe the text in clusters, by putting sentences similar to each other in the same cluster, then selecting the highest score sentence from each cluster.
- Statistical-based approach: It uses statistical analysis to determine the most important words and sentences by word frequency and TF-IDF.
Word frequency: This method score sentence depending on the frequency of each word after removing stop words, this method is the most used for sentence scoring.
TF-IDF: It calculates the word's importance in the document, and it is better than word frequency in text summarization.
- Semantic-based approach:
This method represents the structure of the text using RST, RST is used to know the rhetorical structure between text graphs and build a summary by selecting important text graphs identified by RST.

Abstractive text summarization techniques:

- Sequence to sequence: This is a deep learning method that consists of two parts: encoder and decoder.
 1. First, Input text and reference summary will be tokenized.
 2. Then, they will be encoded by the encoder.
Encoder: It converts the sequence of words to a sequence of vectors by using convolutional, recurrent, or transformer neural network.
 3. Then, the vector will be passed to the decoder.
Decoder: it uses the output vector from the encoder to predict the summery sequence on a token per token basis.

Table 1: Comparison between different Arabic text summarization algorithms

<u>Paper Name</u>	<u>Year</u>	<u>Method</u>	<u>Performance (F-measure)</u>
Multi-Document Arabic Summarization Using Text Clustering to Reduce Redundancy	2014	Clustering	60%
Hybrid-based Arabic Single-Document Text Summarization Approach Using Genetic Algorithms	2016	Hybrid-based	54.76%
Multi-document extractive text summarization based on firefly algorithm	2019	Firefly Algorithm	57.52%
Arabic Text Summarization Using AraBERT Model Using Extractive Text Summarization Approach	2020	AraBERT Model	54%
Extractive Arabic Text Summarization Using Modified PageRank Algorithm	2020	PageRank	67.99%

Evaluation measures of summary:

- Text Quality Measures: There are several ways to calculate text quality:
 - Grammatically: summary shouldn't contain incorrect words.
 - Coherence and structure: the structure of the summary must be good.
 - Non-redundancy: summary shouldn't contain redundant information.
 - Reference clarity: The summary must be clear.
- Co-Selection measures:
 - Precision, Recall, and F-score.
 - Relative Utility.

Similar systems for text summarization:

- Lakhasly: website summarize Arabic and English text.
- Paraphrasetool: website summarize Arabic and English and many other languages.
- RESOOMER: website allow summarizing and analysis of text.
- Quillbot: website summarize documents and papers to key points.

3- Analysis and Design

We will discuss the best technique that we found to summarize Arabic text which is Modified PageRank algorithm.

3.1 System Overview

3.1.1 System Architecture

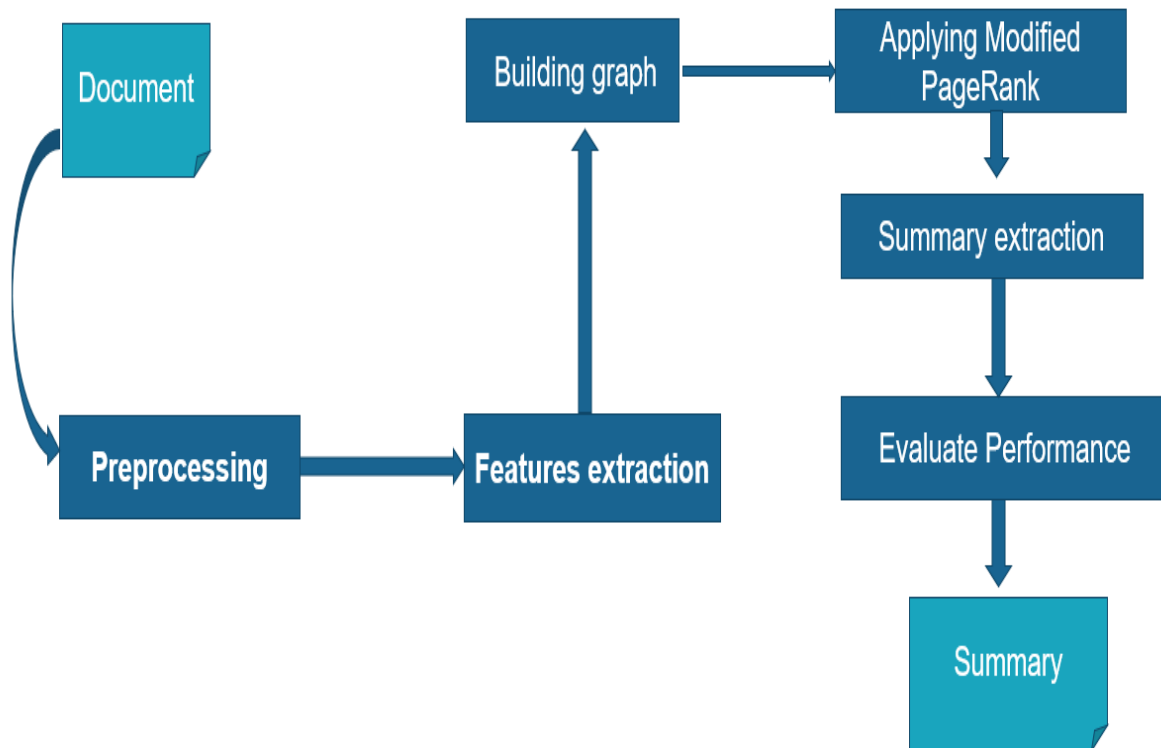


Figure 2. System Architecture

1. Preprocessing:
 1. We split text into sentences then words
 2. Remove stop words
 3. Apply stemming algorithm
 4. Apply morphological analyzer
2. Features extraction:
 1. Calculate cosine similarity
 2. Calculate counts of nouns

3. Build graph: represent the relation between sentences where the vertices represent the sentences Every two vertices are connected with an edge, which has a weight equal to cosine similarity
4. Apply Modified PageRank: in this step, each sentence is given a special rank that expresses the import of the sentence
5. Summary Extraction: nodes are sorted depending on their final rank, and add sentences to the summary with check the overlapping between the sentences if then the overlapping is very high then, this sentence is neglected to prevent redundancy.
6. Evaluate Performance: Calculate accuracy

3.1.2 System Users

A. Intended Users:

Arabic script readers:

There are more than 300 million Arabic speakers in the world and with the massive increase of online documents it is hard to read all of that, our system makes it easier for the user and saves his time by showing him a summary that contains the important sentences in documents

B. User Characteristics

- Can use the internet and websites
- Can copy and paste text from website or document to our website

3.2 System Analysis & Design

3.2.1 Use Case Diagram

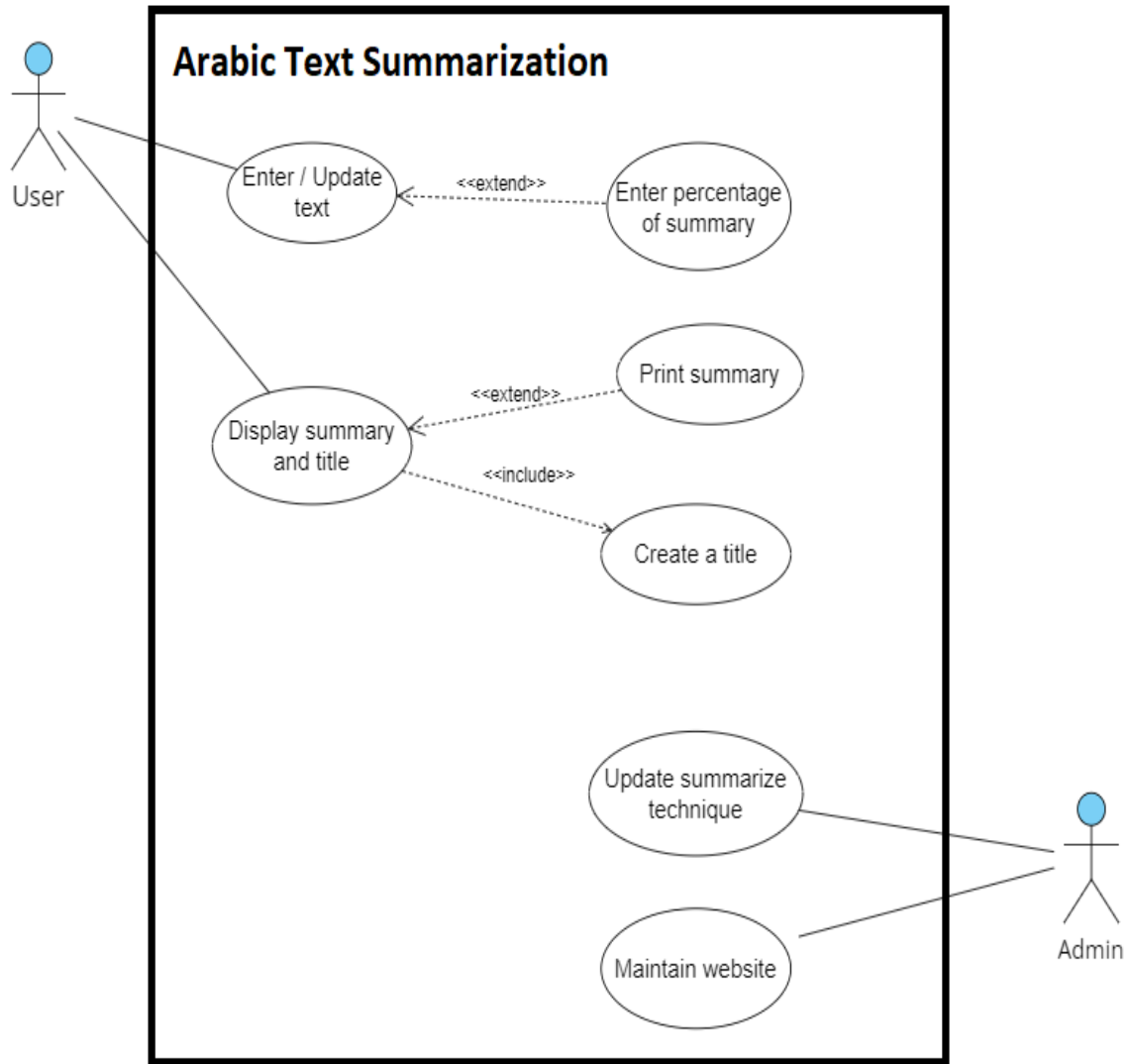


Figure 3. Use Case diagram

1. Enter / Update text: user can enter or update the text to summarize
2. Enter percentage: user can choose the percentage of the summarizer from the original text
3. Display summary and title: display summary text and title of the text

4. Print summary: user can print summary text (it is an option)
5. Create title: create a title to the text
6. Update summarizer technique: admin can update the technique we used to improve performance
7. Maintain website: admin can update and add features and improve performance

3.2.2 Class Diagram

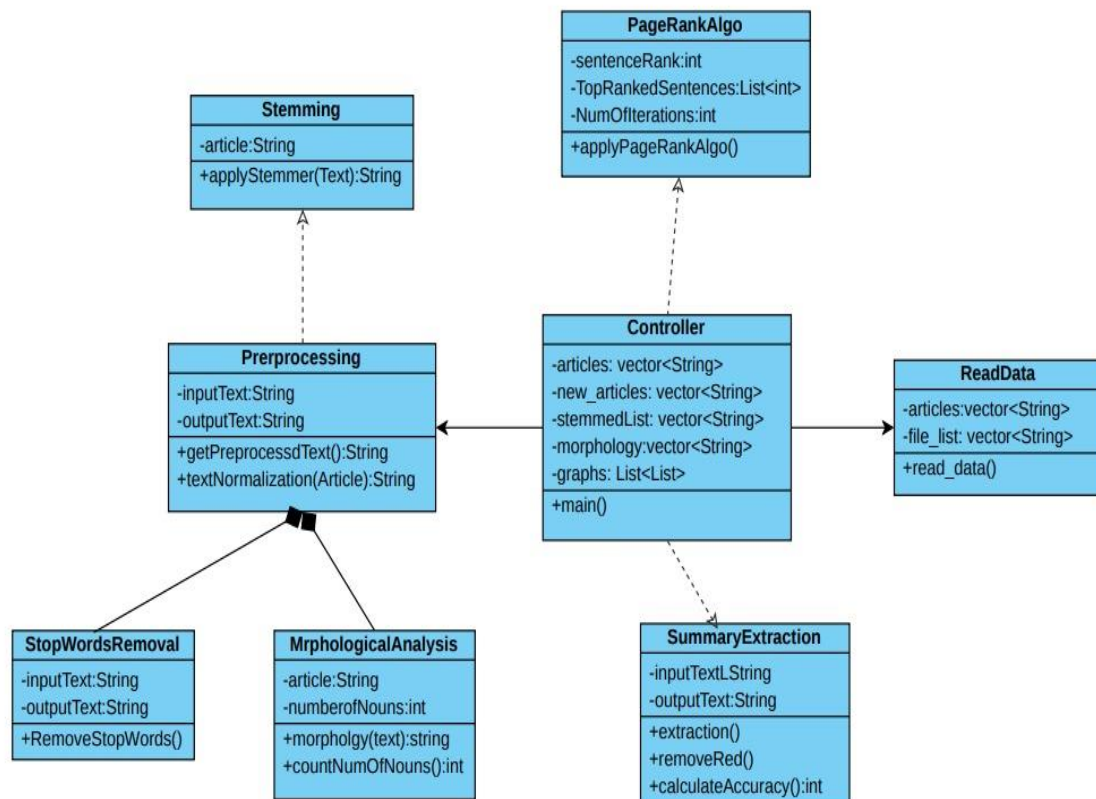


Figure 4. Class diagram

1. Read Data: read articles from the dataset.
2. Preprocessing: Apply text normalization and sentence segmentation, then use stop removing word class.

3. Stop Word Removal: remove stop words from the text.
4. Stemming: take a word and extract the root of the word to reduce the number of distinct words in the document to make a better term frequency calculation.
5. Morphological Analysis: Map each word in the text if it is a verb or noun then count number of nouns in each sentence.
6. PageRank Algo: Uses graph and number of nouns to give initial rank to sentences then update rank for 1000 iterations.
7. Summary Extraction: arrange sentences by rank then remove sentences with the lowest rank then arrasentencesence according to its place in text.
8. Controller: Link classes with each other.

3.2.3 Sequence Diagram

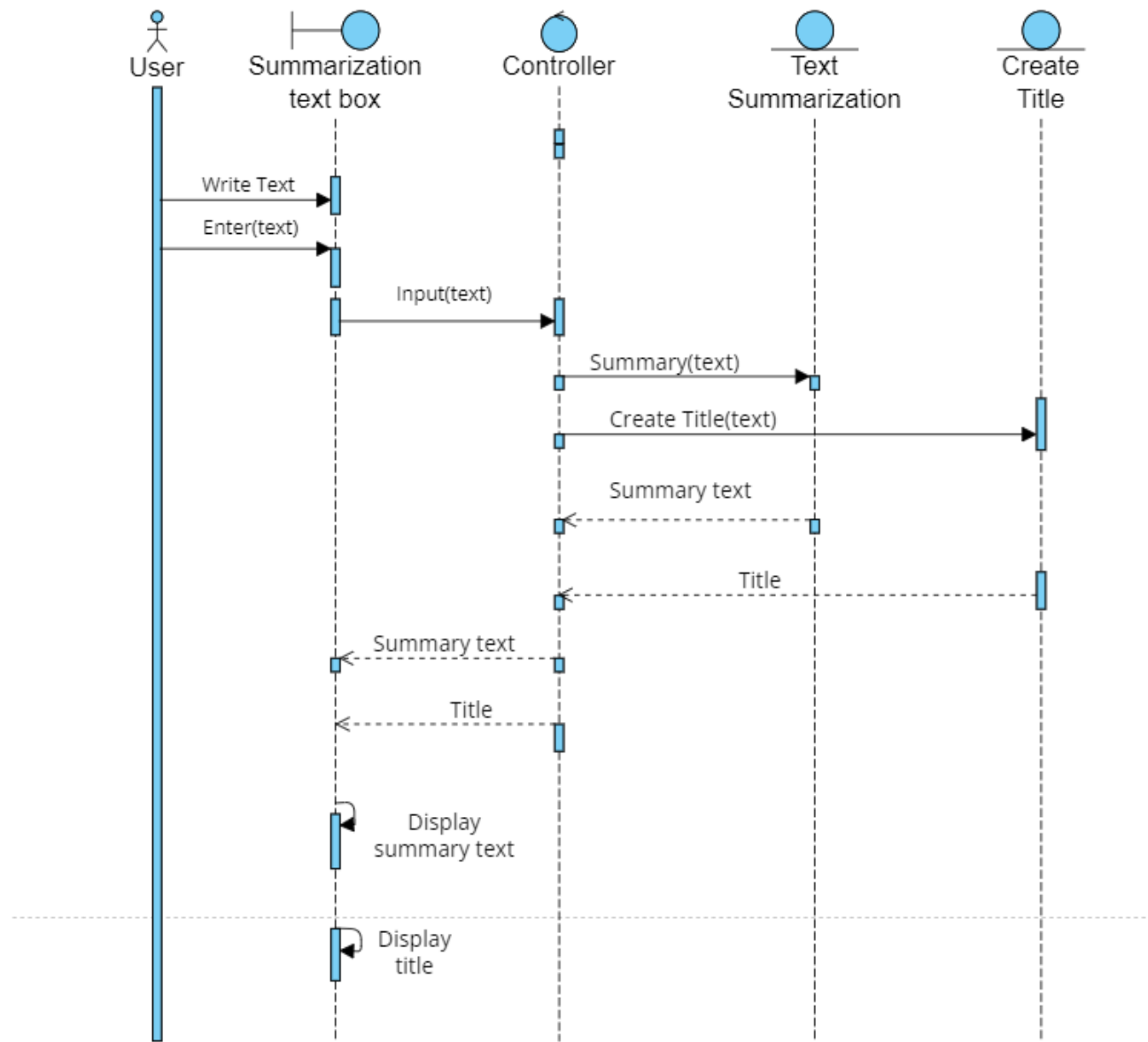


Figure 5. Sequence diagram

3.2.4 Database Diagram

No database was needed for this project.

4- Implementation and Testing

1. Preprocessing

This module is concerned with preparing the text. we start by reading the text, then apply Normalization, in this step punctuations, digits and diacritics are eliminated from the sentence, also the letters are restored to their original form in Arabic language. Then, Tokenization which means split text to word and save in list. Then, Take a list of words, remove all stop words in text and finally apply ISRIS Stemmer to extract the root of word to reduce the number of distinct words in the text and find similar words.

Function: Preprocessing

Input: Text

Output: list of words

```
Read text
Normalization (text)
words ← tokenization (text)
for word in words:
    if word in stop word list then
        remove word
    else
        word = Stem (word)
end if
```

Preprocessing Algorithm 1

2. Morphological analysis:

Every word in the sentence takes a tag representing its Part of Speech (POS) position in the sentence like (verb, noun, preposition, etc....) and this help us identify the important words. then we determine count of nouns words in each sentence in text as it will be used as initial rank for modified page rank algorithm.

Function: Morphological-Analysis

Input: Text

Output: list of nouns

```
for each word in text
    lemmas← lemmatize-text (WORD) //get word position
for each sentence in sentences
    count←0
    for each word in sentence
        if lemmas(word)='noun'
            count++           //count number of nouns
```

Morphology Algorithm 1

3. Building graph:

In this step the text is represented as a graph and the vertices of this graph is sentences and edge between sentences is weighted edge which represents cosine similarity between sentences.

The input of this function is the text after stemming.

We should calculate TF-IDF for the Sentences.

$$TF = \frac{\text{Number of occurrences of term in document}}{\text{Number of all terms in text}}$$

$$IDF = \log \frac{\text{Number of sentences in text}}{\text{Number of sentences containing the term}}$$

$$TF-IDF = TF * IDF$$

Then calculate cosine similarity between each two sentence and this will be the weighted edge between these two sentences.

$$\text{Cosine Similarity (s1, s2)} = \frac{\sum \text{TF-IDF}(s1) * \text{TF-IDF}(s2)}{\sqrt{\sum \text{TF-IDF}(s1)^2} * \sqrt{\sum \text{TF-IDF}(s2)^2}}$$

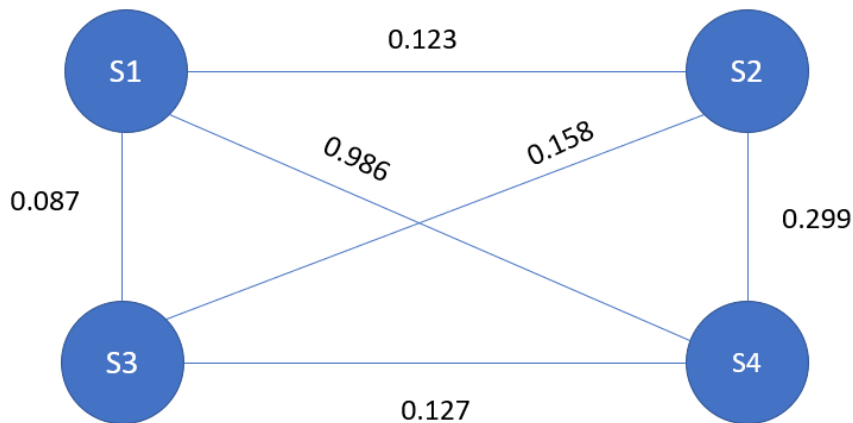


Figure 6. Graph

Function: Graph

Input: Stemmed Text

Output: Graph

```

for I =0 to NumberOfSentences
  for j =1 to NumberOfSentences
    if I=0
      AddVertex (I) //vertex is index of each sentence
      tfidf=TF-IDF (stemmedtext[I], stemmedtext[j])
      cosineSimilarity = cosine_similarity(tfidf[0], tfidf[1])
      add_edge(i,j,cosineSimilarity)
  //add adge between two sentences and weight=cosineSimilarity
  
```

Building Graph Algorithm 1

4. Modified page rank algorithm:

First, give every sentence initial rank equal its own the number of nouns.

Then Apply PageRank algorithm with number of iterations 1000 to update rank of every sentences.

Mpr: New rank for sentence in each iteration.

$E(g, v_i)$: edge between sentence and each sentence.

$PR(v_i)$: Previous rank of sentence.

N: number of sentences in document.

d: constant number=0.85

$$mpr(g) = (1 - d) \times \sum_{i=1}^N \frac{PR(v_i) \times E(g, v_i)}{N - 1}$$

Function: PageRank Algorithm

Input: Number of nouns and Graph

Output: list of ranking

Initial-rank=numberofnouns

d=0.85

count= numberOfSentences

graphs-rank = []

For iterations=0 to 1000

 for indx1=0 to count

 sum = 0

 for indx2 =1 to count

 if iterations=0

 sum =sum+(graph[indx2] * intial_rank [indx2])

Modified PageRank Algorithm 1

5. Summary Extraction:

Sort sentences descending because the sentence that has biggest ranking it is most important and so we are arranged sentences in order of importance.

Then, remove sentences with ranking less than 0.4 and Check cosine similarity between sentences and remove sentence of cosine similarity more than 0.7

Function: Summary Extraction

Input: Sentences and Graph

Output: Summary

```
sorted the sentences based on ranking
for s in sentences:
    for g in graph:
        if cosine similarity between s, g > 0.7
            remove s
        end if
    end for
end for
```

Summary Extraction Algorithm 1

6. Dataset:

The dataset used for evaluation is The Essex Arabic Summaries Corpus (EASC) It consists of 153 documents. Each document has five corresponding summaries written by human experts. EASC includes 10 subjects: art, music, environment, politics, sports, health, finance, science, and technology, tourism, religion, and education.

7. Testing:

The evaluation Matrices that are used to evaluate summary are precision, recall and F-measure. These matrices evaluate summary of the system for the dataset

document against five corresponding summaries written by human experts in the dataset.

Precision: This metric the correct text size that is returned by the system.

$$\text{precision} = \frac{\text{Extracted Summary} \cap \text{Provided Summary}}{\text{Extracted Summary}}$$

Recall: This metric coverage system reflects the ratio of the extracted relevant sentences.

$$\text{Recall} = \frac{\text{Extracted Summary} \cap \text{Provided Summary}}{\text{Provided Summary}}$$

F-measure: This metric Makes a balanced relation among recall metric and precision metric.

$$F - \text{measure} = \frac{2 * \text{percision} * \text{Recall}}{\text{Percision} * \text{Recall}}$$

Function: Calculate Accuracy

Input: Summary Text

Output: F-measure

Calculate precision
Calculate Recall
Calculate F-measure
Display F-measure

Calculate Accuracy Algorithm 1

8. Evaluation results:

Rouge is essentially a set of metrics for evaluating automatic summarization text. It works by comparing an automatically produced summary against set of reference summaries (typically human-produced).

Rouge-N: measures unigram, bigram, trigram and higher order n-gram overlap.

Rouge-L: measures longest matching sequence of words.

F-Measure:

- Rouge 1= 60%
- Rouge 2= 49%
- Rouge 3=59%

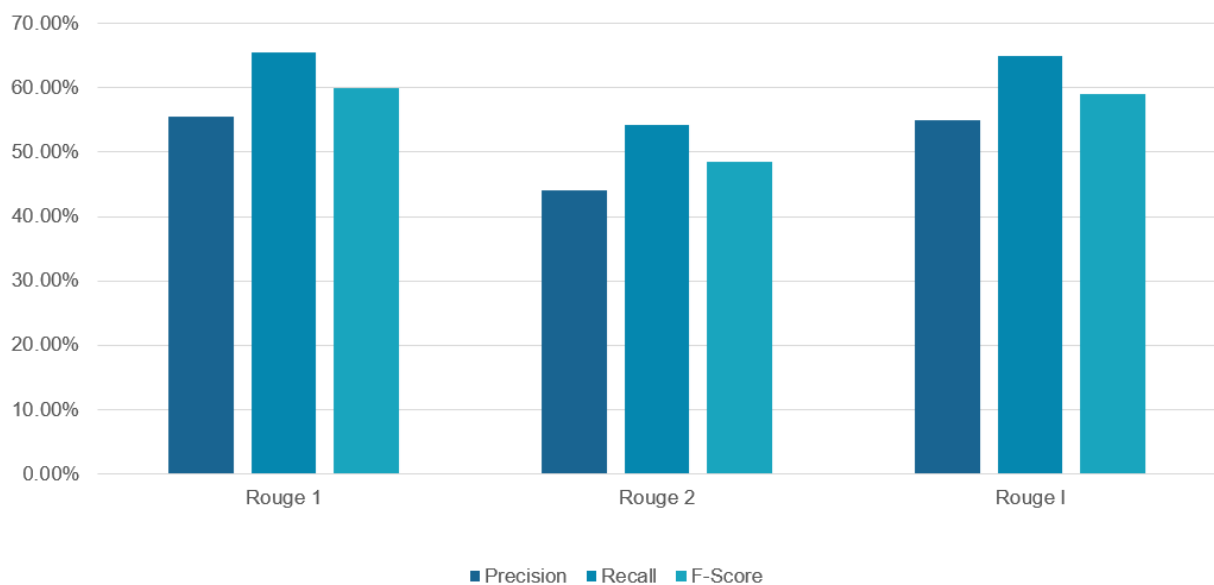


Figure 7. Evaluation results

9. Extract Title

Calculate the phrase score for each key phrase according to the equation:

$$PF_{Score} = \left(\frac{1}{Ph_Pos + 1} \right) + T_Thresh + \sum_{i=1}^{Ph_Len} TF + (Ph_Dist) + \log_2 PhF + (PhF \times IDF)$$

Function: Extract Title

Input: Arabic Textual Document

Output: List of the Extracted Keyphrases

```
BEGIN
  WHILE Not EOF
    Remove Arabic Stop Words
    Stem Arabic Text
    Compute Term Frequency (TF) of each Arabic
    Identify each Paragraph P in the document
    WHILE NOT END of (P)
      Identify each Phrase Ph in the document
      Compute Phrase Frequency (PF)
      IF (PF) > 1
        Extract Phrase (Ph) attributes
        Compute Phrase score (Pscore)
        Save P, Ph, PF, and Pscore into (Phrases-List)
      END IF
    END WHILE
    WHILE NOT END of Phrases-List
      IF PF > 1
        Choose the highest frequency phrase
      END IF
      IF Ph is a Substring from any phrase in Phrases-list
        Remove Ph from the Phrases-List
      END IF
    END WHILE
  ENDWHILE
  Rank candidate phrases Ph in Phrases-List in descending order according to their PFScore
```

Extract Title Algorithm 2

10.UI Design:

- UI design is implemented using local website.
- Tools used for website:
 - ❖ For font end:
 - Html:describe the structre of the web page.Its elements tell the browser how to display the content.
 - CSS: describes how html elements displayed on screen.it controls the layout of the webpage.
 - JavaScript: is a scripting language that enables us to create dynamically updating content, control multimedia, animate images, and pretty much everything else.
 - Bootstrap:is the most popular CSS framework.
 - ❖ To connect frontend and backend:
 - Flask

5- User Manual

To run local website:

Installation Guide:

Install anaconda and python

From anaconda prompt run these commands:

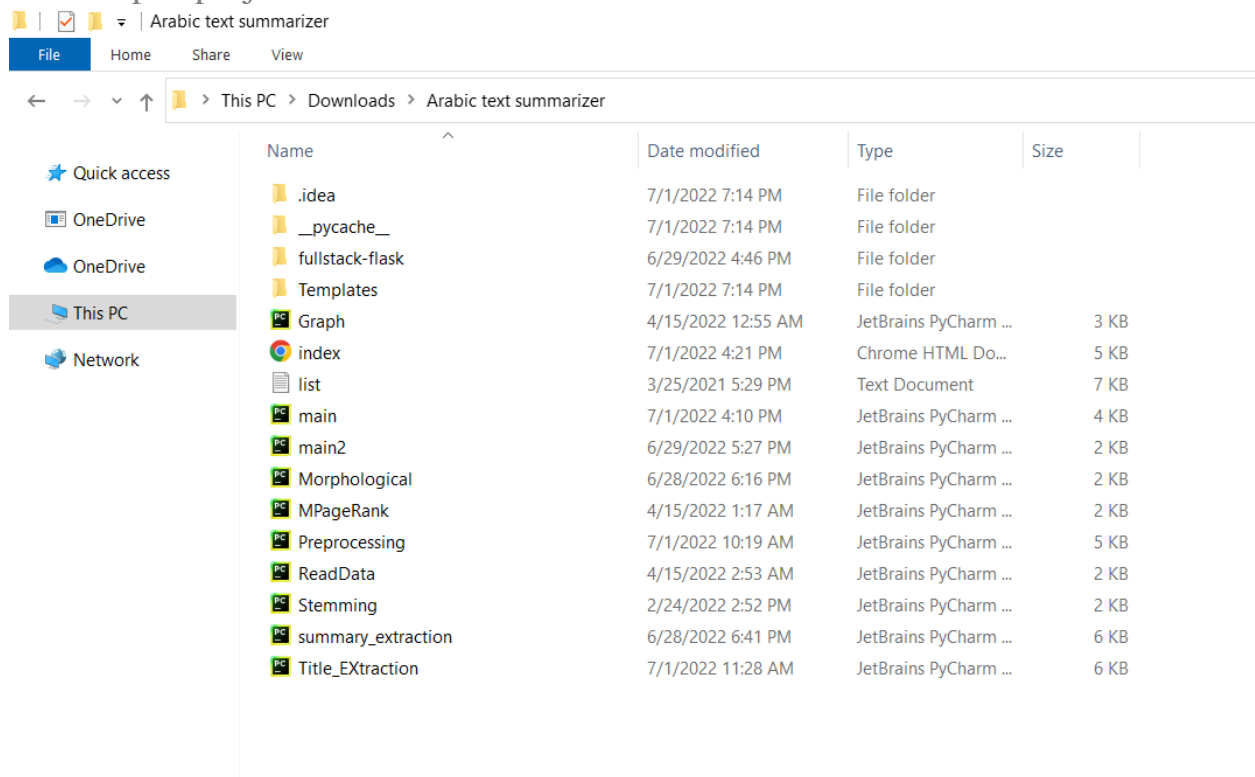
```
pip install jyserver
```

```
pip install qalsadi
```

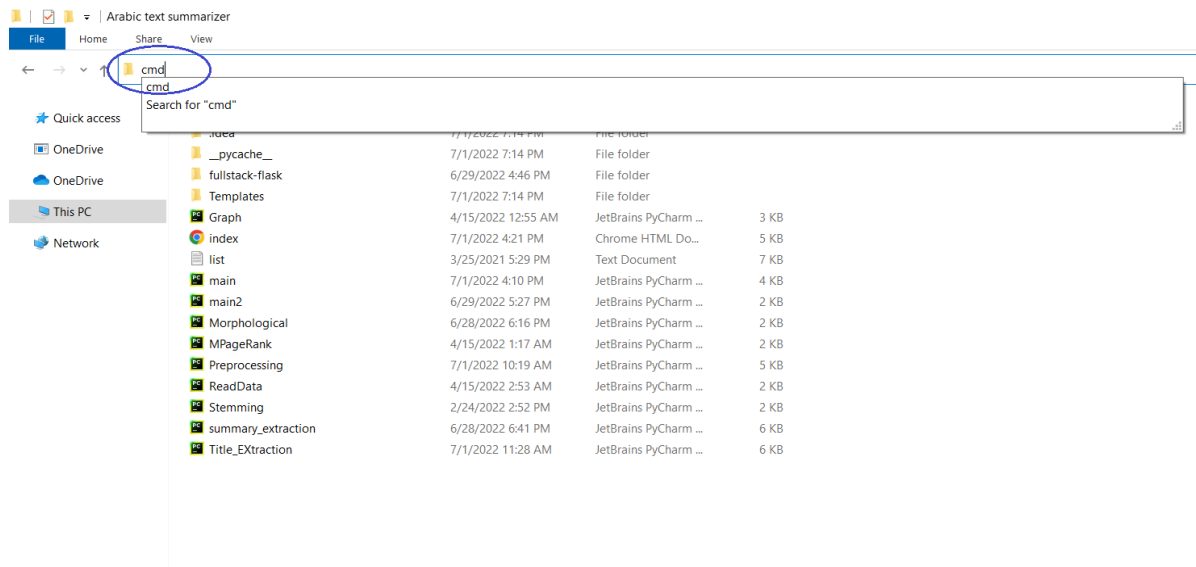
```
pip install self
```

```
pip install flask
```

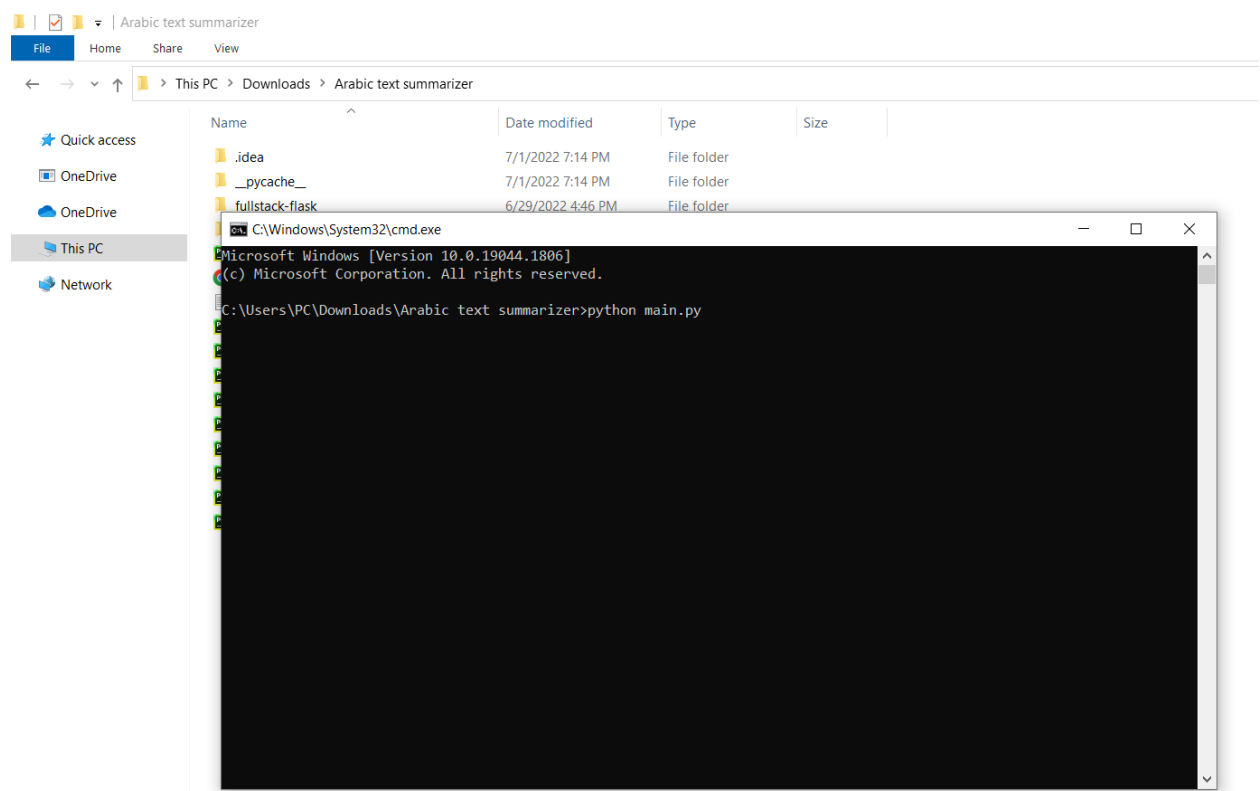
1. Open project file



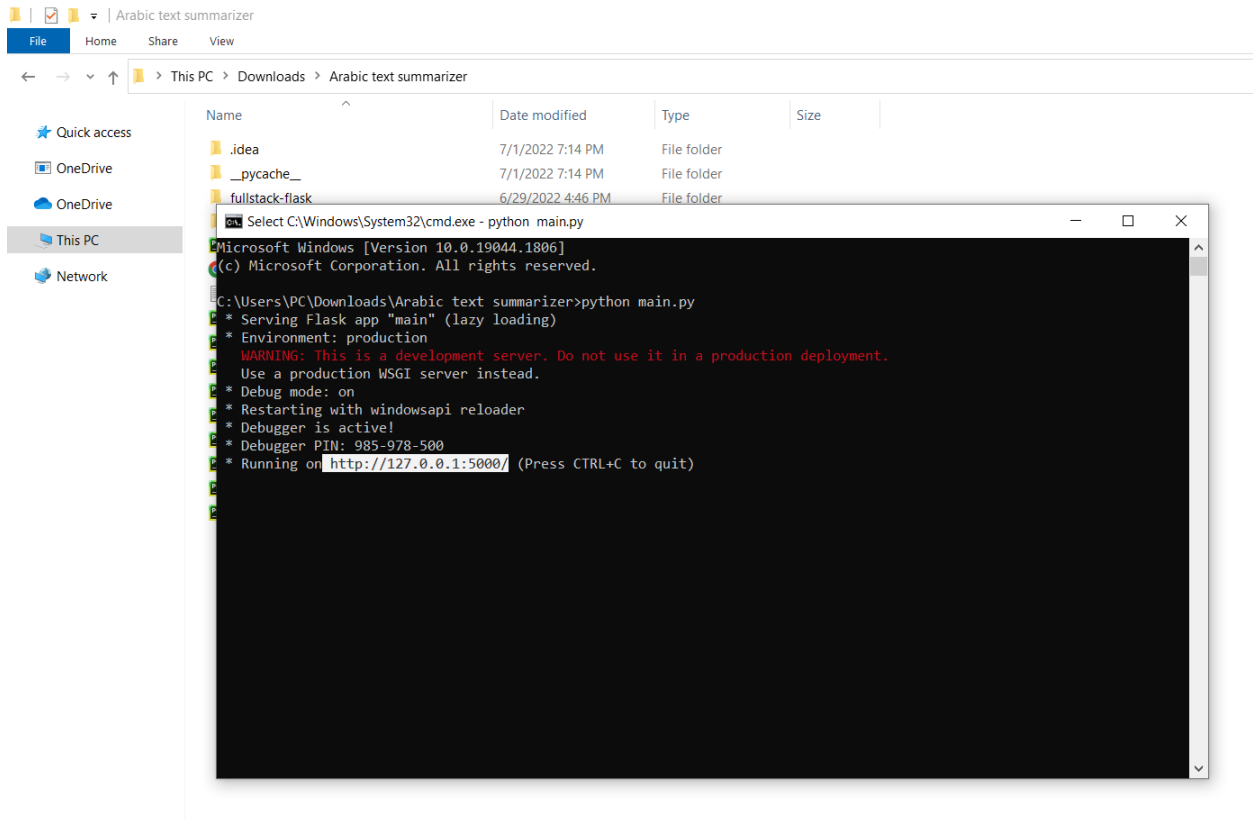
2. Replace path of project with cmd and press enter



3. Write in cmd python main.py and press enter



4. Copy address that will appear



5. Paste address on browser and insert text you want title for in the input text textbox and click create title button to extract title or summarize to extract summary.

Summarized Output

لودفيج فان بيتهوفن مؤلف موسيقى ألماني ولد عام 1770 م في مدينة بون. يعتبر من أبرز عباقرة الموسيقى في جميع العصور، وأبدع أعمالاً موسيقية خالدة. له الفضل الأعظم في تطوير الموسيقى الكلاسيكية. قدم أول عمل موسيقي وعمره 8 سنوات.

تشمل مؤلفاته للأوركسترا تسعة سيمفونيات وخمس مقطوعات موسيقية على البيانو ومقطوعة على الكمان. كما ألف العديد من المقطوعات الموسيقية كمقدمات للأوبرا. من أجمل أعماله السمفونية الخامسة والسادسة والتاسعة. ظهر تميزه الموسيقي منذ صغره، فنشرت أولى أعماله وهو في الثانية عشر من عمره عام 1783 م. اتسعت شهرته كعازف بيانو في سن مبكرة، ثم زاد إنتاجه وذاع صيته كمؤلف موسيقى. عانى بيتهوفن كثيراً في حياته، عالياً وصحياً، فبالرغم من أن أباه هو معلمه الأول الذي وجه اهتمامه للموسيقى ولقنه العزف على البيانو والكمان، إلا أنه لم يكن الأب المثالي، فقد كان مدمناً للكحول، كما أن والدته توفيت وهو في السابعة عشر من عمره بعد صراع طويل مع المرض، تاركة له مسؤولية العائلة. مما منعه من إتمام خطته والسفر إلى فيينا، عاصمة الموسيقى في ذلك العصر. فهل كان التأليف الموسيقي هو نوع من أنواع العلاج والتغلب على المشاكل بالنسبة لبيتهوفن. ولكن بيتهوفن، صاحب الألحان واجه بعض الخلافات مع معلمه، وعندما سافر هايدن إلى لندن، تحول بيتهوفن إلى معلمين آخرين مثل سالييري وشينك وألبريشنبرجر. وحاول أن يشق لنفسه طريق كعازف في عاصمة الموسيقى، وسرعان ما لاقى مكانة كبرى خاصة في الأوساط الأرستقراطية، وبالفعل مارالت أعماله حتى اليوم من أهم ما أنتجته الموسيقى الكلاسيكية.

Input Text

لودفيج فان بيتهوفن مؤلف موسيقى ألماني ولد عام 1770 م في مدينة بون. يعتبر من أبرز عباقرة الموسيقى في جميع العصور، وأبدع أعمالاً موسيقية خالدة. له الفضل الأعظم في تطوير الموسيقى الكلاسيكية. قدم أول عمل موسيقي وعمره 8 سنوات.

تشمل مؤلفاته للأوركسترا تسعة سيمفونيات وخمس مقطوعات موسيقية على البيانو ومقطوعة على الكمان. كما ألف العديد من المقطوعات الموسيقية كمقدمات للأوبرا.

بدأ بيتهوفن يفقد سمعه في الثلاثينيات من عمره إلا أن ذلك لم يؤثر على إنتاجه الذي ازداد في تلك الفترة وتميز بالإبداع. من أجمل أعماله السمفونية الخامسة والسادسة والتاسعة. وقد توفي في فيينا عام 1827 م.

شهدت مدينة بون الألمانية ميلاد الفنان العبقري لودفيج فان بيتهوفن في 16 ديسمبر عام 1770، وتم تعميده في 17 ديسمبر 1770. ظهر تميزه الموسيقي منذ صغره، فنشرت أولى أعماله وهو في الثانية عشر من عمره عام 1783 م. اتسعت شهرته كعازف بيانو في سن مبكرة، ثم زاد إنتاجه وذاع صيته كمؤلف موسيقى. عانى بيتهوفن كثيراً في حياته، عالياً وصحياً، فبالرغم من أن أباه هو معلمه الأول الذي وجه اهتمامه للموسيقى ولقنه العزف على البيانو والكمان، إلا أنه لم يكن الأب المثالي، فقد كان مدمناً للكحول، كما أن والدته توفيت وهو في السابعة عشر من عمره بعد صراع طويل مع المرض، تاركة له مسؤولية العائلة. مما منعه من إتمام خطته والسفر إلى فيينا، عاصمة الموسيقى في ذلك العصر. فهل كان التأليف الموسيقي هو نوع من أنواع العلاج والتغلب على المشاكل بالنسبة

Summarize

Create Title

لودفيج بيتهوفن

6- Conclusion and Future Work

6.1 Conclusion

the Arabic language is spoken by more 300 million people all over the world, and despite that, it suffers from poor researches in it, because it suffers from many problems and challenges. Therefore, The Arabic language is considered one of the strongest and most difficult languages.

Arabic text summarization is of great importance, especially in last period with the massive increase of online documents it become hard to read all of that. Our system saves time for users and make it easier for them, as it displays the important sentences and points in document. One of the most important points of this project is that it supports the Arabic language.

This research tries to enhance the performance of the generated summaries by applying the Modified PageRank algorithm. Morphological analyzer is used to overcome the problems of Arabic structure complexity and to extract nouns to use in the process of building the graph, and cosine similarity was used to weigh the edges between sentences. A Modified PageRank algorithm was used to extract the summary, this algorithm was used by making the initial rank of the sentence as the number of nouns it has, and the weight of the edge is the cosine similarity between the connected nodes.

The process of summarization starts by reading the documents, then normalizing data, removing stop words, stemming, morphological analyzer then finally applying the graph and getting the summary. EASC is used as a standard corpus in the testing stage. According to the results, the Modified PageRank returns better results when the number of iterations used is equal to 1000. Page Rank algorithm returns better results than the methods created before it so we use it and our final F-measure is 60.

6.2 Future Work

- Summarizing the videos and audio recording, then display text containing the most important sentences that were said in them.
- Translate the summary text to any other language.
- Add a voice to system to read the summary text.

References

- ❖ [1] Hani Alghanem , Rashan H Ajamiah .(December 2020). Arabic text summarization approaches: A comparison study
- ❖ [2] Lamees MahmoudAl Qassem , DiWang , ZaidAl Mahmoud, HassanBarada, AhmadAl-Rubaie, Nawaf I.Almoosa. (5 November 2017.). Automatic Arabic summarization: A survey of methodologies and systems
- ❖ [3] Reda Elbarougya, Gamal Beherya, Akram El Khatibb.(July 2020). Extractive Arabic Text Summarization Using Modified PageRank Algorithm.
- ❖ [4] Qasem A. Al-Radaideh , Dareen Q. Bataineh, (March 2018). A Hybrid Approach for Arabic Text Summarization Using Domain Knowledge and Genetic Algorithms.
- ❖ [5] [Said Desouki](#) , [Molham Al-Maleh](#),([December 2020](#)). Arabic text summarization using deep learning approach.
- ❖ [6]Hassan M.Najadat,Ismail I.Hmeidi,Mohammed N.AL-Kabi,Maysa Mahmoud Bany Issa,(Febraury 2016).Automatic KeyPhrase Extractor from Arabic Documents.
- ❖ [7]Aqil Azmi, Nouf I. Altmami,(November 2018). An abstractive Arabic text summarizer with user-controlled granularity.