##### A Software Engineering Project

##### On

TEXT ANALYZER AND SUMMARIZER

as the partial fulfillment of

**Semester -VI study**

for the degree of

**Bachelor of Science (Computer Science)**

## (Affiliated to Veer Narmad South Gujarat University)

during the academic year 2021-2022.

**Developed By: Guided By:**

Master Zufra Dr.Yesha Mehta



Sarvajanik Education Society

## Shree Ramkrishna Institute of Computer Education and Applied Sciences

## Behind P.T Science College, M.T.B.College Campus, Athwalines, Surat-395 001.



### 

Sarvajanik Education Society

## Shree Ramkrishna Institute of Computer Education and Applied Sciences

# Behind P.T. Science College, M.T.B College Campus, Athwalines, Surat-395 001.

### CERTIFICATE

## DEPARTMENT OF COMPUTER SCIENCE

This is to certify that

Student name Exam no

1. Master Zufra 2184

has / have successfully completed his / her / their software engineering project work entitled

TEXT ANALYZER AND SUMMARIZER

as the partial fulfillment of

**Semester -VI study**

for the degree of

**Bachelor of Science (Computer Science)**

## (Affiliated to Veer Narmad South Gujarat University)

during the academic year 2021-2022.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

##### Internal Project Guide HOD

#### Dept. of Computer Science

##### Date : \_\_\_\_\_\_\_\_\_\_\_\_ Place : Surat

### University Examination:

**Examination No. : \_\_\_\_\_\_\_\_\_\_\_\_**

**Signature of External Examiners : \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**INDEX for Project Documentation**

|  |  |  |  |
| --- | --- | --- | --- |
| **SR.NO** | **CONTENTS** | | **Page. No** |
| **1.** | **Introduction** | | **1** |
|  |  | Project Profile |  |
| **2.** | **System Introduction** | | **2** |
|  |  | System Definition, Objective and scope |  |
|  |  | Hardware & Software Requirements |  |
| **3.** | **Requirement Analysis & Modeling** | | **4** |
|  |  | Expected working of system |  |
|  | Data Flow Diagrams (DFD) |  |
|  | Process Specification |  |
|  |  | Data Dictionary |  |
| **4.** | **Design** | | **15** |
|  |  | ER Diagram & Database Design |  |
|  | System Architecture |  |
|  | System Flowchart |  |
|  | Form & Report Designs (Screenshots) Layouts |  |
| **5.** | **Coding** | | **19** |
|  |  | Coding Approach/style |  |
|  | Code Snippets |  |
| **6.** | **New Tools/ Technologies learned/used** | | **23** |
| **7.** | **Bibliography & References** | | **24** |

1. **PROJECT INTRODUCTION**
   1. **PROJECT PROFILE:**

**PROJECT TITLE:** TEXT ANALYZER AND SUMMARIZER

**DEVELOPED BY:** Zufra Master (TY CS DIV-A)

**INSTITUTE:** Shree Ramkrishna Institute Of Computer Application and Applied Science

**PROJECT DEFINITION:** As the project title suggests, Analyzing a text involves breaking down its ideas and structure to understand it better, think critically about it, and draw conclusions and Text summarizer helps in summarizing the text. The main purpose is to provide reliable summaries of web pages or uploaded files depends on the user’s choice. The unnecessary sentences will be discarded to obtain the most important sentences.

**TOOLS:**

**FRONT-END:** Python

**BACK-END:** Python, MySQL

**SCRIPTING – LANGUAGE:** HTML ,CSS , JavaScript

**PROJECT GUIDE:** **Dr. Yesha Mehta**

1. **SYSTEM INTRODUCTION**
   1. **SYSTEM DEFINITION AND OBJECTIVE**
      1. **SYSTEM DEFINITION:**

* Encryption helps us by protecting data from hackers. In network communication, the same techniques can be used in saving passwords.
* The project is based on NLP.NLTK(Natural language toolkit) is the next level library used for performing Natural language tasks like Sentiment Analysis, removing stopwords, tokenization, stemming, part of speech tagging, word cloud, Entity Extraction, etc.
* To analyze means to break something down into its parts and examine them. Analyzing is a vital skill for successful readers.
* In text summarizer, this library is used to remove stop words in English vocabulary and to convert these words to root forms.
  + 1. **OBJECTIVE:**
* Password encryption aims at hiding the content of the message by translating it into a code
* The goal of this summarizer is to help people for reducing their time.
* People can get rid from reading long and toxic text. Instead they can do only focus on important lines.
* This will help you anticipate how a text will develop an idea and improve your own writing.
* Learn how summarizing differs from quoting and paraphrasing, as well as how to apply summarizing to check your understanding of what you read and help you remember the main ideas.
  1. **HARDWARE & SOFTWARE REQUIREMENT:**

**HARDWARE:**

* Processor: Intel core i3-3110M
* Processor Speed: 2.40 GHz
* Hard Disk: 1 TB
* Main Memory: 4 GB

**SOFTWARE:**

* Operating System: Windows 10 Pro
* Browser: Any
* Programming Tools: PYCHARM , VS CODE, PHPMYADMIN
* Scripting Languages: HTML , CSS , BOOTSTRAP, JAVASCRIPT
* Documentation Tool: Microsoft Word

1. **REQUIREMENT ANALYSIS AND MODELING**

**3.1 EXPECTED WORKING OF SYSTEM:**

* **Registration Module :**  Those who visit this site for the first time, must first register and find a registration form which they must fill out.
* **Login Module:** Only registered users are able to log in via email and password on this site.
* **Database Module:** People who have registered their data in the registration form will be stored in the database and will also appear on the home page.
* **Password Encryption Module :** Those who have registered their data in the sign-up form their password is encrypted and stored in the database.
* **Analyze and Summary Module:** For those who want to analyze and summarize their text, they can write their text in the text box. They can able to analyze and summarize the statement, they can read analyzed and summarized statement.
* **POS Tagging:** Part-of-speech (POS) tagging is a popular Natural Language Processing process which refers to categorizing words in a text (corpus) in correspondence with a particular part of speech, depending on the definition of the word and its context.
* **Sentiment Analysis:** Sentiment analysis is contextual mining of text which identifies and extracts subjective information in source material, and helping a business to understand the social sentiment of their brand, product or service while monitoring online conversations.
* **Entity Extraction:** Entity extraction models recognize specific data in text that you target based on your business needs.
* **Stemming:** Stemming is the process of reducing a word to its word stem that affixes to suffixes and prefixes or to the roots of words known as a lemma.
* **Word Cloud:** Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance.
* **Frequency Distribution:** A frequency distribution records the number of times each outcome of an experiment has occurred.

**3.2 DATA FLOW DIAGRAM (DFD):**

* **Context level DFD:**

1. **Password Encryption :**

Cipher Password

(Encrypted Data)

Plain Password

(Data to Encrypt)

1. **Text Analyzer and Summarizer :**

Summary

Text

1. **Entity Extraction:**

TEXT

Extracted Entity

1. **Sentiment Analysis:**

Text

Training the data

* **First Level DFD For Password Encryption :**

Database Table

Encrypt

Generate Password

Plain Password

* **First Level DFD For Analyzing:**

Result

Login

Stemming

Word Tokenization

POS tagging

Sentence Tokenization

Valid Or not

UserEmail & Password

Passwor

No

Yes

Removing Stopwords

Frequency Distribution

* **First Level DFD For Summarization:**

Login

UserEmail & Password

Passwor

Valid Or Not

No

Yes

Summarized Text

Sentence Filtering

Key\_phrase Extraction

Summarization

Process

* **First Level DFD For Entity Extraction:**

Sentence Segmentation

Raw Text

Machine Learning Based Model

Event Extraction

Lexicon Based Model

Rule Based Model

Entity Recognition

Parts Of Speech tagging

Tokenization

* **First Level DFD For Sentiment Analysis:**

Positive, Negative or Neutral

Searching Keywords

Raw Text

Sentiment

Identification

Feature

Selection

Sentiment Classification

Sentiment

Score

**3.3 PROCESS SPECIFICATION:**

* + **Process 1.0 For Password Encryption:**

|  |  |
| --- | --- |
| Name : | Password Encryption |
| Description : | In this process, once registered, the user’s password is encoded and stored in the database. |

* + **Process 1.0 For Analyzing:**

|  |  |
| --- | --- |
| Name : | Login |
| Description : | This process will store the logon user e-mail and password. |

* + **Process 2.0 For Analyzing:**

|  |  |
| --- | --- |
| Name : | Write Text |
| Description : | This process will store the paragraph of text you would like to analyze. |

* + **Process 3.0 For Analyzing:**

|  |  |
| --- | --- |
| Name : | Text Mining |
| Description : | This process will analyze the text and carry out the following operation : tokenization, stopwords, lexicon normalization, POS tagging. |

* + **Process 4.0 For Analyzing:**

|  |  |
| --- | --- |
| Name : | Result |
| Description : | This process will display the result of the operations performed in text mining. |

* + **Process 1.0 For Summarizing:**

|  |  |
| --- | --- |
| Name : | Login |
| Description : | This process will store the login to the user’s e-mail and password. |

* + **Process 2.0 For Summarizing:**

|  |  |
| --- | --- |
| Name : | Write Text |
| Description : | This process will store the paragraph of the texts to be summarized. |

* + **Process 3.0 For Summarizing:**

|  |  |
| --- | --- |
| Name : | Summarizing |
| Description : | This process will summarize your previously added text. |

* + **Process 4.0 For Summarizing:**

|  |  |
| --- | --- |
| Name : | Summarized Text |
| Description : | This process will display the summarized text. |

* + **Process 1.0 For Entity Extraction:**

|  |  |
| --- | --- |
| Name : | Login |
| Description : | This process will store the login to the user’s e-mail and password. |

* + **Process 2.0 For Entity Extraction:**

|  |  |
| --- | --- |
| Name : | Write Text |
| Description : | This process will store the paragraph of the texts to be summarized. |

* + **Process 3.0 For Entity Extraction:**

|  |  |
| --- | --- |
| Name : | Entity recognition |
| Description : | This process helps us to sort the unstructured data and detect the important information. |

* + **Process 4.0 For Entity Extraction:**

|  |  |
| --- | --- |
| Name : | Entity Extraction |
| Description : | This process will display the extracted text. |

* **User Registration Database:**

|  |  |
| --- | --- |
| **Name:** | User Registration |
| **Alias:** | None |
| **Where used/How used:** | * User details(Input) * User login(Output) |
| **Supplementary Information:** | Id, Name, Email, Department, Designation, Password |

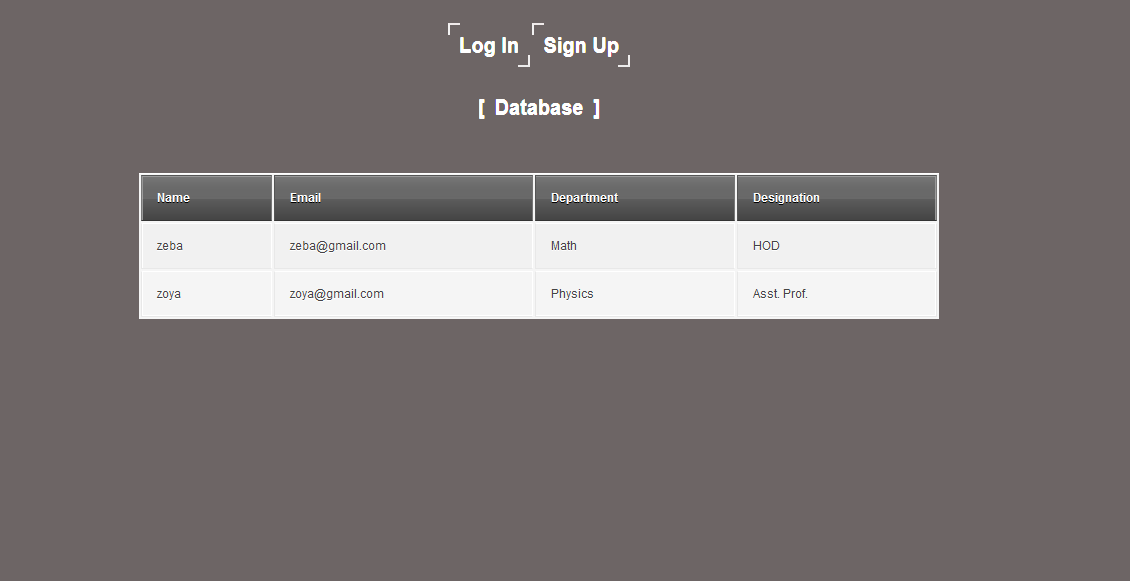
* **User Login Database:**

|  |  |
| --- | --- |
| **Name:** | User Login |
| **Alias:** | None |
| **Where used/How used:** | * User login details(Input) * Successfully login(Output) |
| **Supplementary Information:** | Id, Password, Email |

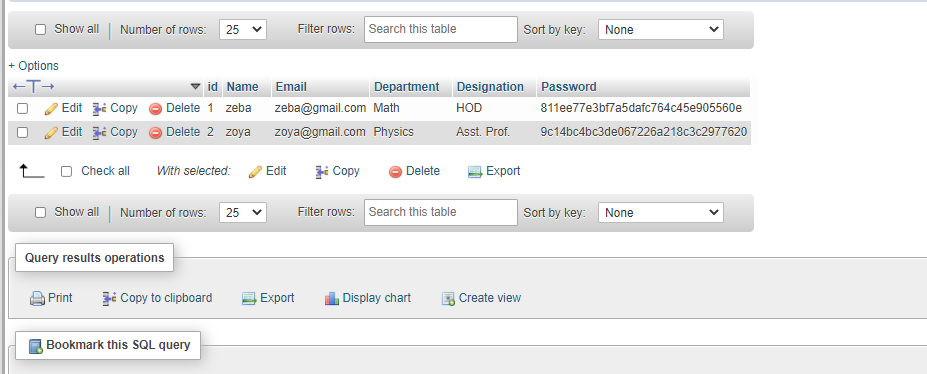
**4. DESIGN :**

**FORM & WEB-PAGES DESIGN (SCREENSHOT) LAYOUT**

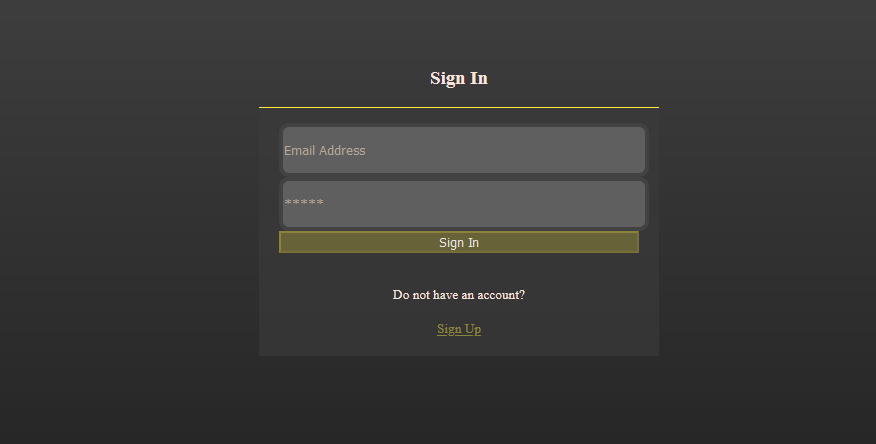
* + **Home Page :**



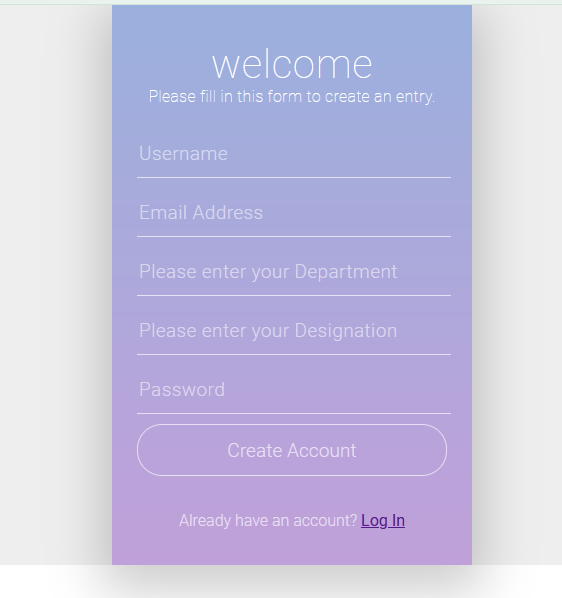
* + **Database :**



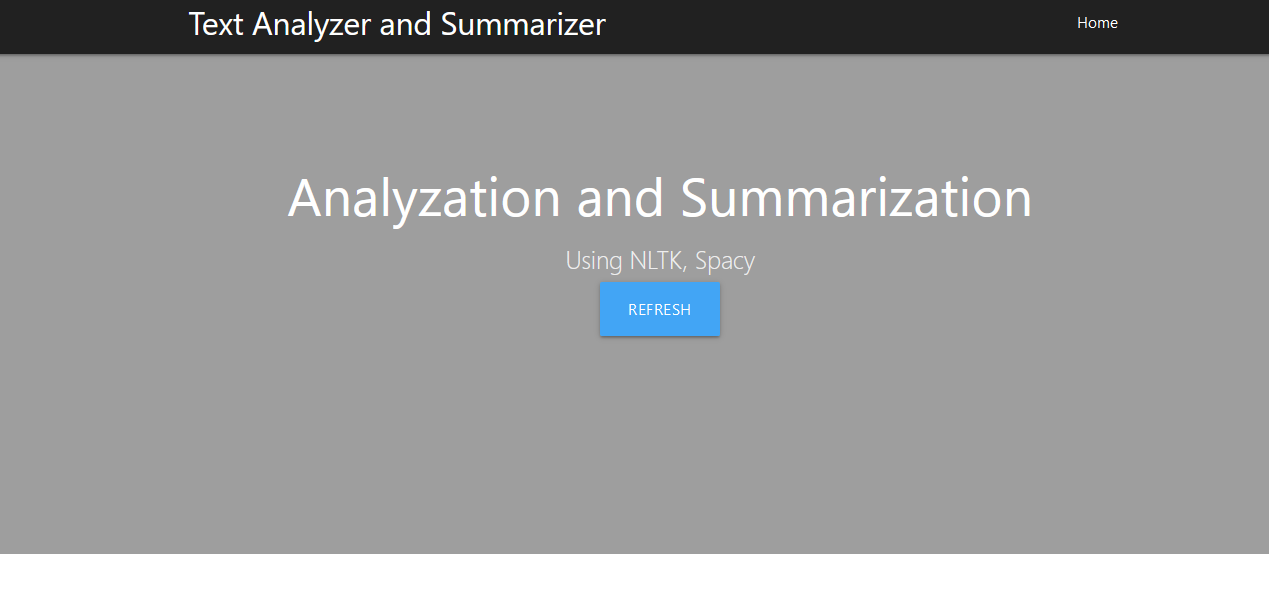
* + **Login Page :**

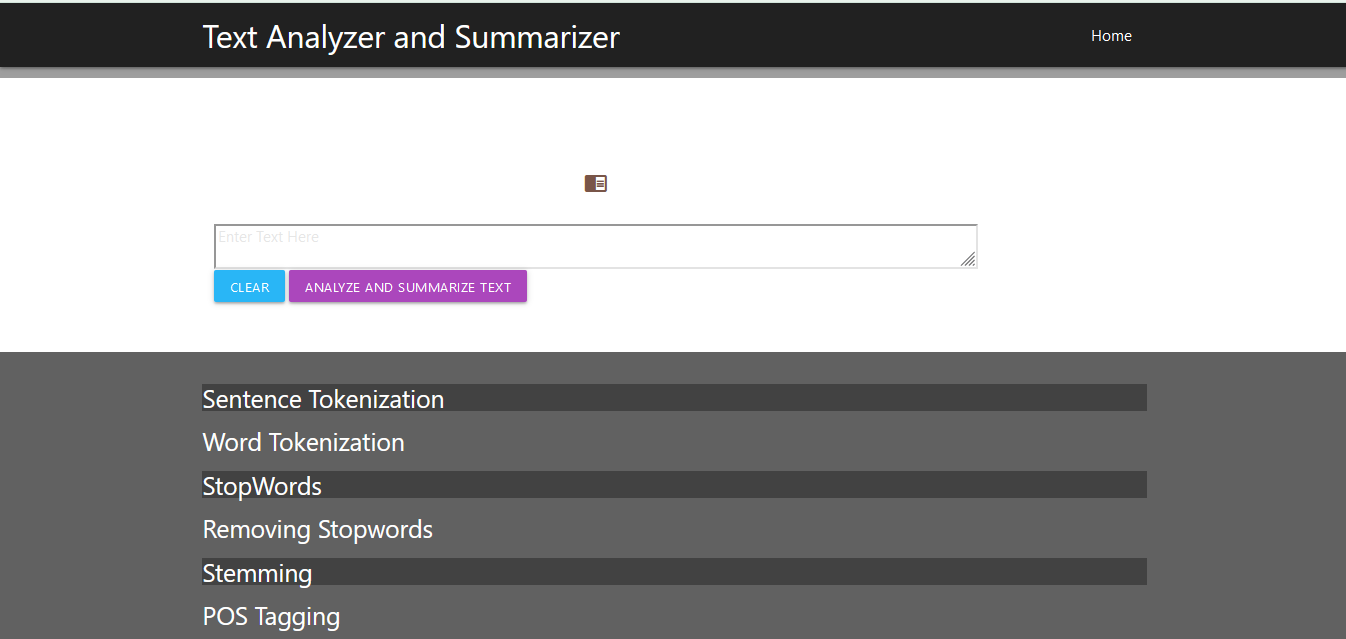


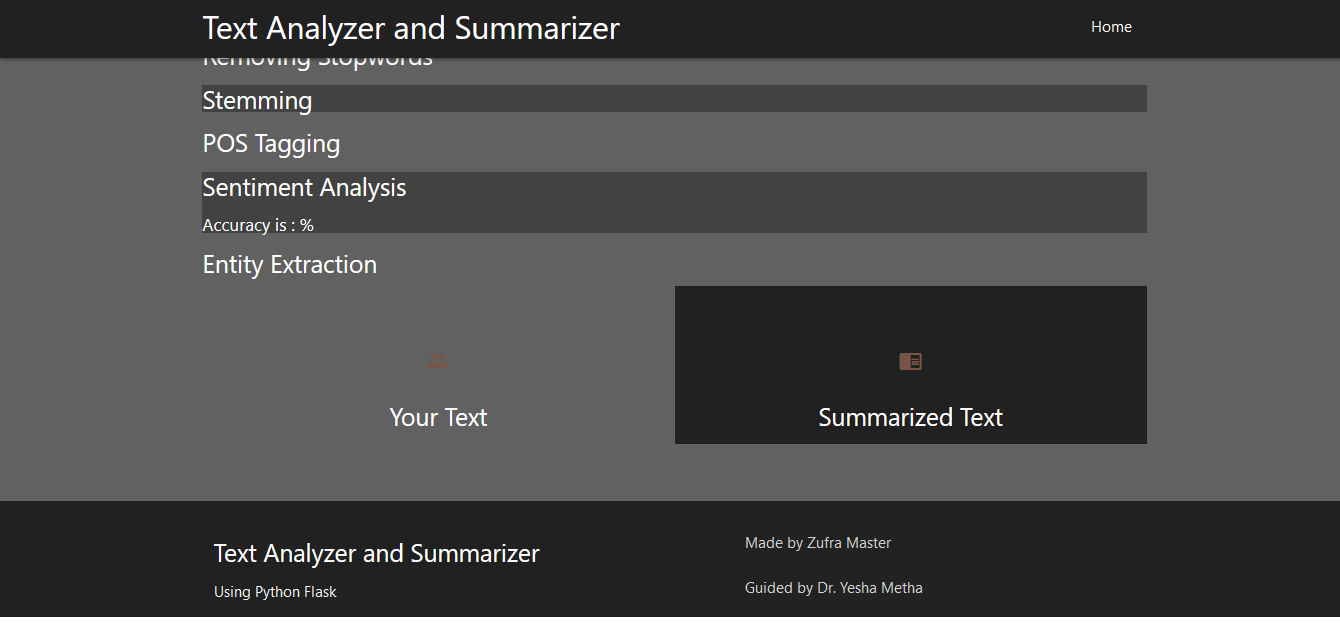
* **Registration Page :**

****

* + **Text Analyzer and Summarizer Page :**







**5. CODING :**

* + **Home Page:**

<html>

<body>

<table>

        <thead>

            <tr>

                <th>Name</th>

                <th>Email</th>

                <th>Department</th>

                <th>Designation</th>

            </tr>

        </thead>

        <tbody>

            <tr>

                <td>{{ Name}}</td>

                <td>{{ Email}}</td>

                <td>{{ Department}}</td>

                <td>{{ Designation}}</td>

            </tr>

        </tbody>

    </table>

</body>

</html>

* LogIn Page :

 <body>

<h3 >Sign In </h3>

         <input type="email" class="form-control" name="uemail" id="uemail" placeholder="Email Address" required="">

        <input type="Password" name="pwd" id="pwd" class="form-control" placeholder="\*\*\*\*\*" required="">

      <button type="submit name="btnsub" id="btnsub">Sign In</button>

    <p ><small>Do not have an account?</small></p>

     <small>Sign Up</small>

</body>

* Registration Page :

 <div >welcome</div>

      <div >Please fill in this form to create an entry.</div>

        <input type='text' name="uname" placeholder='Username' class='input-line full-width' required></input>

        <input type='email' placeholder="Email Address" name="uemail" class='input-line full-width' required></input>

        <input type="text" placeholder="Please enter your Department" name="udept" class='input-line full-width' required></input>

        <input type="text" placeholder="Please enter your Designation" name="udesig" class='input-line full-width' required></input>

        <input type='password' placeholder='Password' class='input-line full-width' name="pwd" required></input>

      <div><button type="submit>Create Account</button></div>

      <div class='spacing'>Already have an account?

        <span>Log In</span>

* Admin Page :

   <a>Text Analyzer and Summarizer</a>

 <ul >

   <li>Home</li>

 </ul>

   <i class="material-icons">menu</i>

<h2>Analyzation and Summarization</h2>

   <h5> Using NLTK, Spacy </h5>

<h5>Sentence Tokenization</h5>

 <p> tkn\_text </p>

 <h5>Word Tokenization</h5>

 <p>tkn\_word </p>

 <h5>StopWords </h5>

 <p> st\_word </p>

<h5>Removing Stopwords</h5>

<p> fil\_sent </p>

<h5>Stemming</h5>

<p> stem\_word </p>

<h5>POS Tagging</h5>

<p>token</p>

<h5>Sentiment Analysis</h5>

<h4> sasen </h4>

<h5>Entity Extraction</h5>

<div>sresult </div>

<h5 class="center">Your Text</h5>

<p class="light">ctext</p>

<h5 class="center">Summarized Text</h5>

<p class="light"> final\_summary </p>

<h5 class="white-text">Text Analyzer and Summarizer</h5>

  <p class="grey-text text-lighten-4">Using Python Flask</p>

Made by Zufra Master<br><br> Guided by Dr. Yesha Metha

* **Extractive Method :**

**def** home\_page():  
 connection = create\_connection()   
 sql = **"SELECT \* from encryption;"** connection.close()  
 **return** render\_template(**"home.html”**)

sql = **"INSERT INTO encryption(Name, Email, Department, Designation, Password) VALUES (%s,%s,%s,%s,%s); "** val = (  
 request[**"uname"**], request [**"uemail"**], request [**"udept"**], request [**"udesig"**], password)  
connection.close()  
 **return** redirect(**"/"**)

sql = **"SELECT Email,Password FROM encryption WHERE Email = %s AND Password = %s;"** password = request.form[**"pwd"**]  
 val = (request.form[**"uemail"**], password)  
 cursor.execute(sql, val)  
 connection.close()

**def** nltk\_summarizer(rawtext):  
 stopWords = set(stopwords.words(**"english"**))  
 word\_frequencies = {}  
 **for** word **in** nltk.word\_tokenize(rawtext):  
 **if** word **not in** stopWords:  
 **if** word **not in** word\_frequencies.keys():  
 word\_frequencies[word] = 1  
 **else**:  
 word\_frequencies[word] += 1

**def** analyze():  
 **if** request.method == **'POST'**:  
 rawtext = request.form[**'rawtext'**]  
 final\_summary = nltk\_summarizer(rawtext)

**def** admin():  
 **if** request.method == **"POST"**:  
 rawtext = request(**"rawtext"**)  
  
 content = nltk\_summarizer(rawtext)  
 **return** content  
 **return** render\_template(**"admin.html"**)

**6. NEW TOOLS/TECHNOLOGIES LEARNED/USED :**

I) PyCharm and Visual Code – This is the main environment for coding.

II) Database – MySQLite

III) Some CSS pages for design

IV) Template field for passing value

**7. Bibliography & References:**

* **WEBSITE**
* [www.github.com](http://www.github.com)
* [www.youtube.com](http://www.youtube.com)
* [www.w3school.com](http://www.w3school.com)
* [www.tutorialspoint.com](http://www.tutorialspoint.com)