DSCI 560: Group Laboratory 8

Table of Contents

Team Members

1. Hanlu Ma (USC ID: 1392-9443-71)
2. Zhenyu Chen (USC ID: 2242-3773-15)
3. Fariha Sheikh (USC ID: 9625-7343-53)

Installation

Install the require Modules:

1. Install the `sqlalchemy` module:

```
1  pip install sqlalchemy
```

2. Install the `requests` module:

```
1  pip install requests
```

3. Install the `beautifulsoup4` module:

```
1  pip install beautifulsoup4
```

4. Install the `selenium` module:

```
1  pip install selenium
```

5. Install the `webdriver-manager` module:

```
1  pip install webdriver-manager
```

6. Install the `spacy` module:

```
1  pip install spacy
```

7. Install the `numpy` module:

```
1  pip install numpy
```

8. Install the `nltk` module:

```
1  pip install nltk
2
3  # Over your python terminal, do the following:
4  import nltk
5  nltk.download('stopwords')
6  nltk.download('punkt')
```

9. Install the `numpy` module:

```
1  pip install numpy
```

10. Install the `genism` module:

```
1  pip install genism
```

11. Install the `scikit-learn` module:

```
1  pip install scikit-learn
```

12. Install the `matplotlib` module:

```
1   pip install matplotlib
```

Execution

Stage 1: Generate the data

Step 1. Execute the `database` script:

```
1   python3 database.py
2
3   "Please enter the username for the database:" [username] #Enter
    MySQL username.
4   "Please enter the password for the database:" [password] #Enter
    MySQL passowrd.
5   "Please enter the database name:" [database] #Enter the database
    name where you want to store the data.
```

Step 2. Execute the `data_retrieval` script:

```
1   python3 data_retrieval.py
2
3   "Please enter the username for the database:" [username] #Enter
    MySQL username.
4   "Please enter the password for the database:" [password] #Enter
    MySQL passowrd.
5   "Please enter the database name:" [database] #Enter the database
    name where you want to store the data.
```

Stage 2: Run the model and evaluate the performance

Step 3. Execute the `evaluation` script:

Since we directly use the extracted files for analysis, please change the file/directory path in line 18 of `doc2vec` script and line 21 of `word2vec_BoW` script to make sure that the files are located properly.

```
python3 evaluation.py
```