Resume Assignment
CSCE 771 – Fall 2024
Rasoul Nikmehr

The following procedures were implemented using code provided by Dr. Srivastava:

1.   Text extraction from various PDF files using the Tika library.
2.   Visualization of the extracted text content using the WordCloud library.

Three scenarios were analyzed:

Scenario 1: A word cloud was generated for my personal resume. The results (as shown in the following image) highlight:
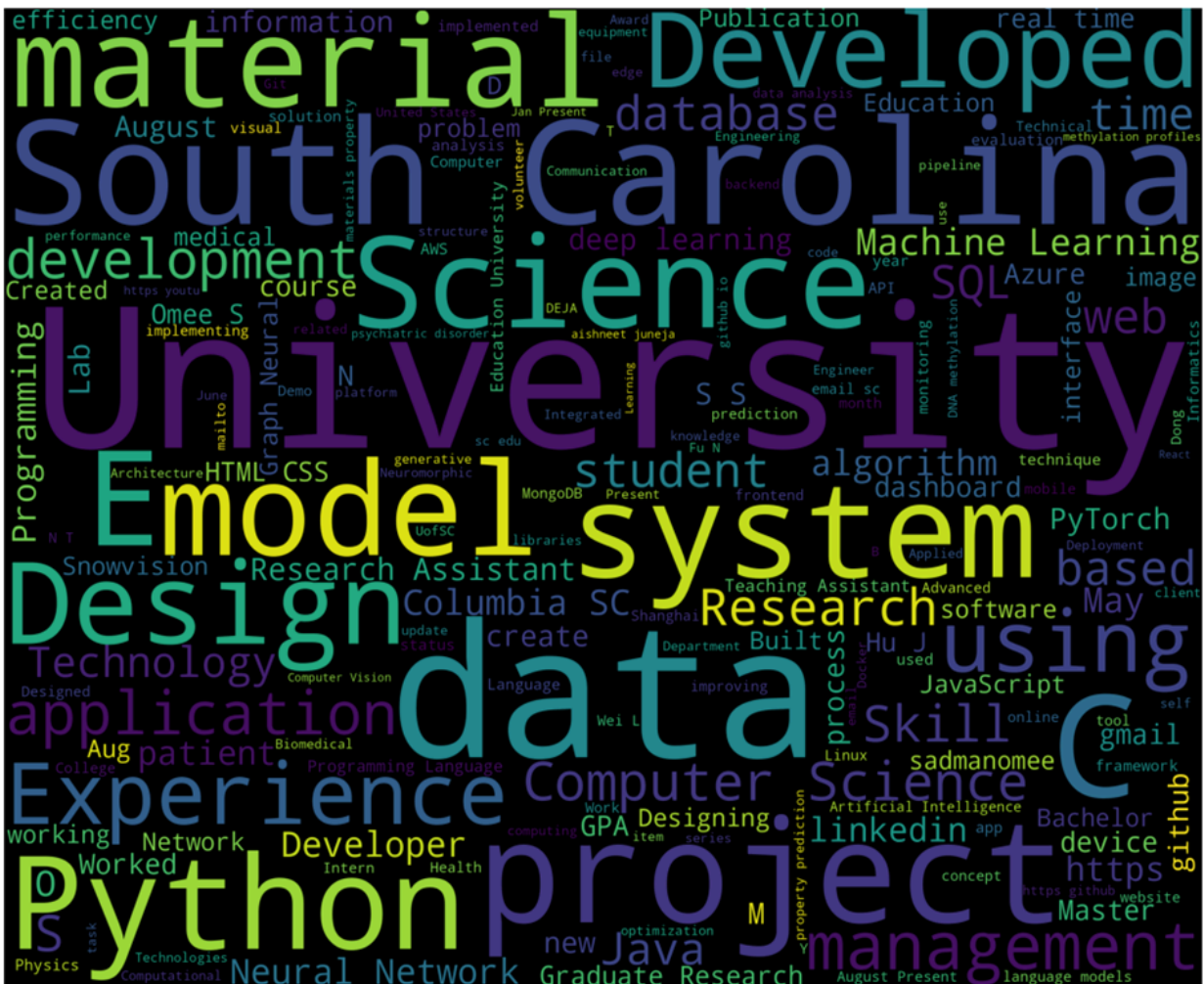
➔ The areas in which I have the most experience or have emphasized as my main interests, such as:

- SQL
- Python
- Database
- Programming
- Java
- Azure
- Design

➔ Words in less bold fonts appear to be more general or those that I have not emphasized as strongly.

Scenario 2: A word cloud was generated for 12 resumes from our current class students. The results (as shown in the following image) indicate:

The dominant words are those most frequently mentioned in resumes. Given that all the students are from the University of South Carolina, it was interesting to see "University of South Carolina" as one of the prominent terms. Other notable words include:

➔ Fields of expertise such as:

● Data
● machine learning
● Python
● Design

➔ Software or applications they have worked with such as:
● JavaScript
● AWS

➔ Their status such as:

● Student
● Graduate student
● Teaching Assistant
● Research (student)

Scenario 3: A word cloud was generated for 30 resumes from a different group of students. The results (as shown in the following image) reveal:

Once again, the dominant words are those most frequently mentioned in the resumes. Like the previous scenario, since all the students are either from the University of South Carolina or have participated in projects there, "University of South Carolina" appeared as a dominant term. Other notable words include:

➔ Fields of expertise or interests and experiences as well as courses taken such as:

- Python
- data science
- Microsoft
- Algorithm
- GitHub
- Java
- Neural Network