CMPE493: Information Retrieval Assignment 3

Elvan Karasu

1 Introduction

In this assignment, PageRank algorithm is implemented using the power iteration method and the Python Standard Library. Names of people in a subset from Reuters-21578 data set is used. [1] The aim is to understand the most popular people from the news articles.

2 Results

```
elvankarasu@Elvans-Laptop pagerank-on-reuters % python main.py  data.txt
Rank   Person Name          Score
---------------------------------
1      nakasone         0.06746033385
2      brodersohn       0.02281149617
3      sumita           0.01740483175
4      papandreou       0.01447786297
5      eyskens          0.01175516759
6      lawson           0.01107373743
7      james-baker      0.01028000185
8      Aqazadeh         0.00991729348
9      reagan           0.00977944106
10     lyng             0.00912776282
11     ciampi           0.00858305241
12     macsharry        0.00827683984
13     Danforth         0.00784585133
14     Graves           0.00767033151
15     de-clercq        0.00750534423
16     goria            0.00732508393
17     Abe              0.00648677056
18     beaumont-dark    0.00596449225
19     fernandez        0.00548762423
20     sprinkel         0.00526954650
```

Figure 1: Top 20 people and PageRank scores

The command for running the program, and the top 20 people with the highest PageRank scores are given in Figure 1. Considering that the data contains news articles from 1987, the results are reasonable. The person with the highest score, Yasuhiro Nakasone, was the prime minister of Japan between 1982-1987. The third person, Andreas Papandreou was the prime minister of Greece between 1981-1989. Other names in the list, such as Eyskens and Goria were also politicians at that time. Therefore, the most popular people found using PageRank algorithm make sense.

References

[1] "UCI Machine Learning Repository: Reuters-21578 Text Categorization Collection Data Set — archive.ics.uci.edu," https://archive.ics.uci.edu/ml/datasets/reuters-21578+text+categorization+collection, [Accessed 04-May-2023].