

CMPE 493 INFORMATION RETRIEVAL

ASSIGNMENT 4:

PageRank for Identifying Central People in News Articles

In this assignment we developed a PageRank-based method to identify the most important people occurring in news articles.

Runs with:

```
python3 page_rank.py data.txt printing_number
```

Interpretation of the Results

Below, you can find the top 50 people and their scores calculated by the algorithm.

When we analyze the years of the news articles, we can see that they are taken from the 1987 archives. The listed names below are mostly the politically important figures of that time. Ronald Reagan was the president of US in 1981-89 period, James Baker is the minister of treasury at the same period. When we analyze the names below, such as Thatcher and Gorbachev, we see they are the prime ministers of UK and SSCB at the same period as well.

In conclusion, we can say that the ranking of page-rank algorithm matches with the importance of the figures that we investigated.

The top 50 people with the highest PageRank scores are found as:

Person: "reagan" Score: 0.06556546567777531
Person: "james-baker" Score: 0.021527490252389127
Person: "nakasone" Score: 0.01632113903740932
Person: "thatcher" Score: 0.014050546940787414
Person: "stoltenberg" Score: 0.01103209303714833
Person: "lyng" Score: 0.011000912628791285
Person: "Weinberger" Score: 0.009692955113532555
Person: "volcker" Score: 0.00964024724810736
Person: "howard-baker" Score: 0.00929597870623089
Person: "yeutter" Score: 0.00875938875346318
Person: "Gorbachev" Score: 0.00815426328472267
Person: "greenspan" Score: 0.008039394047851166
Person: "miyazawa" Score: 0.007676508643099423
Person: "Gephardt" Score: 0.007116481221888682
Person: "kohl" Score: 0.007001932112701154
Person: "poehl" Score: 0.006733806347828709
Person: "shultz" Score: 0.006181068066883479
Person: "Howard" Score: 0.005893298452971033
Person: "ozal" Score: 0.005543999808033283
Person: "chirac" Score: 0.0051599048240646185

Maral Dicle MARAL
22.01.2022

Person: "suharto" Score: 0.005081502515963764
Person: "balladur" Score: 0.00500975239768029
Person: "papandreou" Score: 0.004471853838845383
Person: "lawson" Score: 0.0044680167286993335
Person: "Fitzwater" Score: 0.0043272850377106915
Person: "Abe" Score: 0.004121318210647568
Person: "sumita" Score: 0.004076545575358795
Person: "Mulroney" Score: 0.004046301847072462
Person: "bush" Score: 0.00403225116113239
Person: "Schmidt" Score: 0.003974707047450063
Person: "Roosevelt" Score: 0.0039120184032196475
Person: "Schlesinger" Score: 0.0038092425862221043
Person: "Bentsen" Score: 0.0036859216097523437
Person: "Moore" Score: 0.0036365865392436877
Person: "tamura" Score: 0.00356208855482178
Person: "Ruder" Score: 0.0035398921599758457
Person: "Carter" Score: 0.0035005843419712765
Person: "saddam-hussein" Score: 0.003480497185895697
Person: "honda" Score: 0.0034761398223709496
Person: "Schneider" Score: 0.003467960926974062
Person: "camdessus" Score: 0.0032828897189613707
Person: "Cochran" Score: 0.0032492622787381976
Person: "Yamamoto" Score: 0.003218252400272673
Person: "sprinkel" Score: 0.0031723375855032846
Person: "Leahy" Score: 0.0031597708417689815
Person: "Maekawa" Score: 0.003117445095391914
Person: "john-paul" Score: 0.00310802161137944
Person: "Byrd" Score: 0.0031069697075677217
Person: "Kuranari" Score: 0.0030798084788277954
Person: "keating" Score: 0.003069241412305296

Maral Dicle MARAL
22.01.2022

```
~/Documents/Courses/INFO_RETRIEVAL/Assignment4_LinkAnalysis -- zsh
Last login: Sat Jan 22 14:54:22 on ttys000
(base) maral@eren-MacBook-Pro: Assignment4_LinkAnalysis % python3 page_rank.py data.txt 50
Person: "reagan" Score: 0.0655646567777531
Person: "james-baker" Score: 0.021527490252389127
Person: "nakasone" Score: 0.01632113903740932
Person: "thatcher" Score: 0.014050546940787414
Person: "stoltenberg" Score: 0.01103209303714833
Person: "lyng" Score: 0.011000912628791285
Person: "Weinberger" Score: 0.009692955113532555
Person: "volcker" Score: 0.00964024724810736
Person: "howard-baker" Score: 0.00929597870623089
Person: "yeutter" Score: 0.00875938875346318
Person: "Gorbachev" Score: 0.00815426328472267
Person: "greenspan" Score: 0.008039394047851166
Person: "miyazawa" Score: 0.007676508643099423
Person: "Gephardt" Score: 0.007116481221888682
Person: "kohl" Score: 0.007001932112701154
Person: "poehl" Score: 0.006733806347828709
Person: "shultz" Score: 0.00618106066883479
Person: "Howard" Score: 0.005893298452971033
Person: "ozal" Score: 0.005543999808033283
Person: "chirac" Score: 0.0051599048240646185
Person: "suharto" Score: 0.005081582515963764
Person: "balladur" Score: 0.00500975239768029
Person: "papandreou" Score: 0.004471853838845383
Person: "lawson" Score: 0.004468016728699335
Person: "Fitzwater" Score: 0.0043272850377186915
Person: "Abe" Score: 0.004121318210647568
Person: "sumita" Score: 0.004076545575358795
Person: "Mulroney" Score: 0.004046301847072462
Person: "bush" Score: 0.00403225116113239
Person: "Schmidt" Score: 0.003974707047450063
Person: "Roosevelt" Score: 0.0039120184032196475
Person: "Schlesinger" Score: 0.0038092425862221043
Person: "Bentsen" Score: 0.0036859216097523437
Person: "Moore" Score: 0.0036365865392436877
Person: "tamura" Score: 0.00356208855482178
Person: "Ruder" Score: 0.0035398921599758457
Person: "Carter" Score: 0.0035005843419712765
Person: "saddam-hussein" Score: 0.003480497185895697
Person: "honda" Score: 0.0034761398223709496
Person: "Schneider" Score: 0.003467960926974062
Person: "camdessus" Score: 0.0032828897189613707
Person: "Cochran" Score: 0.0032492622787381976
Person: "Yamamoto" Score: 0.003218252400272673
Person: "sprinkel" Score: 0.0031723375855032846
Person: "Leahy" Score: 0.0031597700417689015
Person: "Maekawa" Score: 0.003117445095391914
Person: "john-paul" Score: 0.00310802161137944
Person: "Byrd" Score: 0.0031069697075677217
Person: "Kuranari" Score: 0.0030798084788277954
Person: "keating" Score: 0.003069241412305296
(base) maral@eren-MacBook-Pro: Assignment4_LinkAnalysis %
```

Figure 1: Screenshot of the page_rank.py working. It prints the highest scored figures on the screen as well as a page_rank_scores.txt file.