

**자연어처리 및 정보검색 02분반 – Homework 1**

Student ID: 20133096

Student Name: 이현재

[Homework1]

|  |
| --- |
| Text Indexing Homework    1) Choose 10 movie scripts from <https://www.imsdb.com/>  (I suggest to make them in TXT files.)  2) Write a code for indexing the texts,  and draw charts for showing the word distributions of them.    \* You can use one of programming language (Python, Java, and C). |

[How to solve]

|  |
| --- |
| 주어진 스크립트들의 모든 단어들의 빈도수를 분석하기 위해, 간단하게 각 단어별로 빈도수를 plotting 하였습니다. 허나, 각 스크립트마다 가지고 있는 다른 단어의 종류수가 매우 많고, 특정단어들은 빈도수가 매우 높으나 대부분의 다른 단어들은 빈도수가 매우 낮았습니다. 따라서, X축은 해당 스크립트의 단어들의 빈도수들을 나타냈고, Y축은 해당 빈도수에 다른 단어들이 몇 개가 있는지를 나타냈습니다. |

[Usage]

|  |
| --- |
| “app.py” 파일을 실행하면, ‘.txt’파일형식의 파일이름을 입력 받습니다. 해당 텍스트파일들은 ‘textfiles’ 폴더 내에 있는 파일 중에 하나입니다. |

[Result]

|  |
| --- |
| 실행하면, 총 3가지 종류의 결과를 출력합니다. 첫번째는 각 단어별로의 빈도수를 출력하고, 두번째는 단어별 빈도수와 해당 빈도수의 단어들 개수를 나타냅니다. 마지막으로 두번째의 값을 표 형식으로 출력합니다.  다음은 Coco.txt 파일을 입력 받았을 때의 3가지 결과입니다. |

[Result-others]

|  |
| --- |
|  |