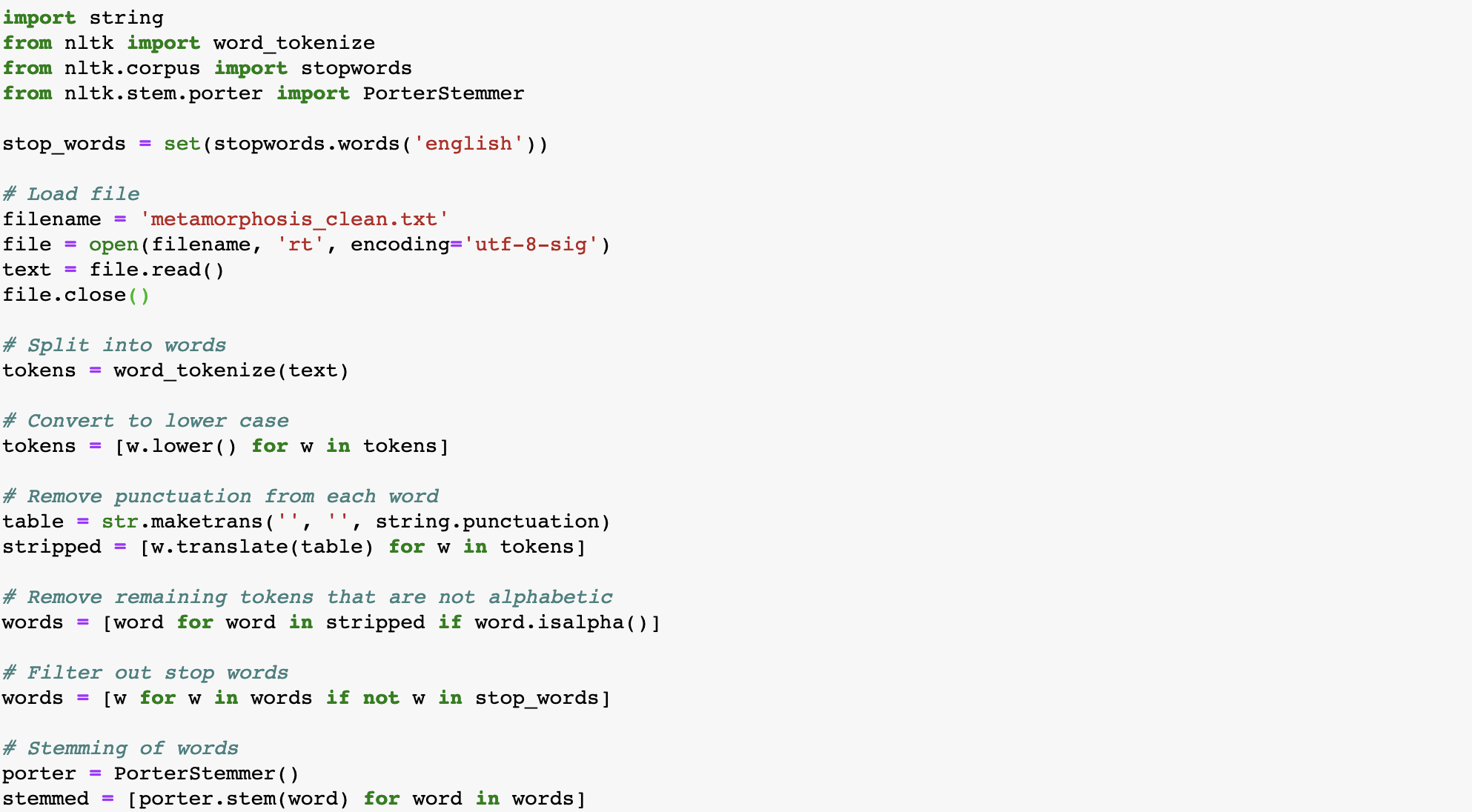
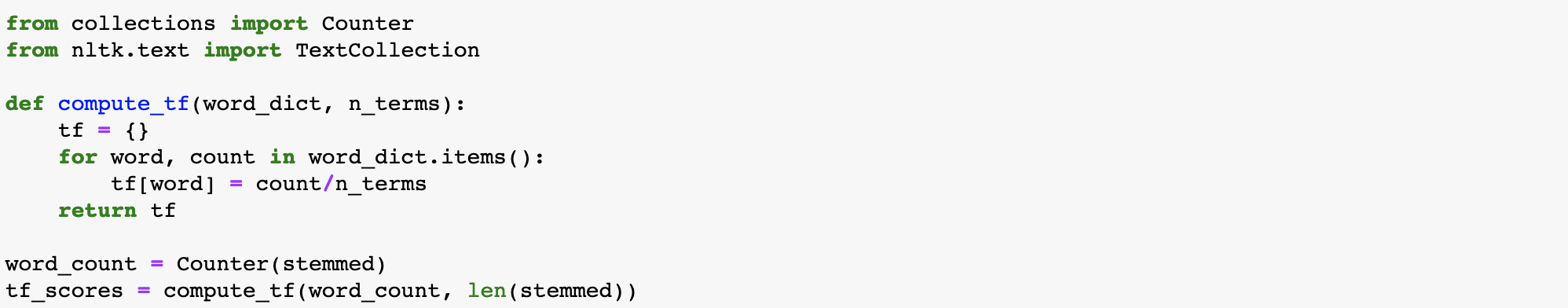
**KSE 643 Homework #3**

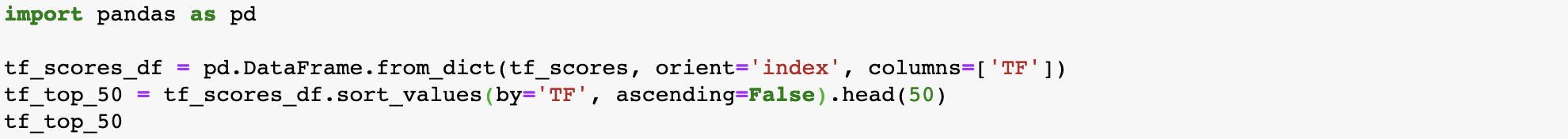


1. For the final set of terms (stems), compute their TF (Term Frequency) scores.



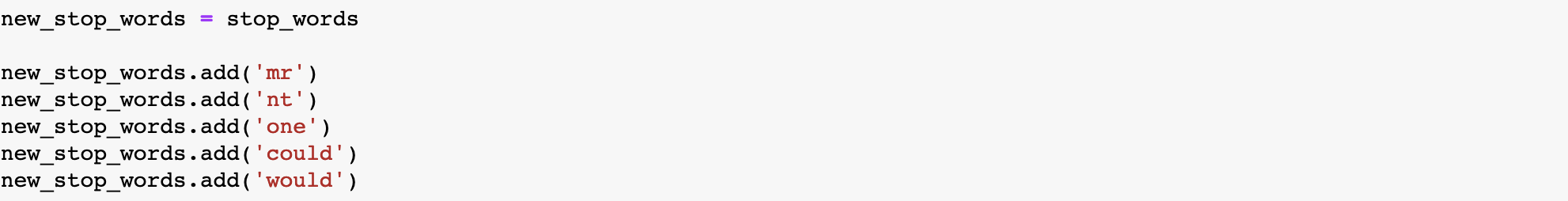
Sample for the first 100 stemmed words: ['one', 'morn', 'gregor', 'samsa', 'woke', 'troubl', 'dream', 'found', 'transform', 'bed', 'horribl', 'vermin', 'lay', 'armourlik', 'back', 'lift', 'head', 'littl', 'could', 'see', 'brown', 'belli', 'slightli', 'dome', 'divid', 'arch', 'stiff', 'section', 'bed', 'hardli', 'abl', 'cover', 'seem', 'readi', 'slide', 'moment', 'mani', 'leg', 'piti', 'thin', 'compar', 'size', 'rest', 'wave', 'helplessli', 'look', 'happen', 'thought', 'nt', 'dream', 'room', 'proper', 'human', 'room', 'although', 'littl', 'small', 'lay', 'peac', 'four', 'familiar', 'wall', 'collect', 'textil', 'sampl', 'lay', 'spread', 'tabl', 'samsa', 'travel', 'salesman', 'hung', 'pictur', 'recent', 'cut', 'illustr', 'magazin', 'hous', 'nice', 'gild', 'frame', 'show', 'ladi', 'fit', 'fur', 'hat', 'fur', 'boa', 'sat', 'upright', 'rais', 'heavi', 'fur', 'muff', 'cover', 'whole', 'lower', 'arm', 'toward', 'viewer', …]

2. Rearrange the terms by their TF scores and print the top 50 words and their respective TF scores in pairs.



|  |  |
| --- | --- |
|  | **TF Scores** |
| **gregor** | 0.029561 |
| **would** | 0.018947 |
| **room** | 0.013193 |
| **could** | 0.012499 |
| **father** | 0.010118 |
| **even** | 0.010118 |
| **sister** | 0.010019 |
| **door** | 0.009622 |
| **mother** | 0.008928 |
| **back** | 0.008233 |
| **time** | 0.007341 |
| **way** | 0.006249 |
| **look** | 0.006051 |
| **nt** | 0.005753 |
| **one** | 0.005753 |
| **open** | 0.005456 |
| **get** | 0.005158 |
| **said** | 0.005059 |
| **littl** | 0.004861 |
| **go** | 0.004861 |
| **still** | 0.004464 |
| **want** | 0.004365 |
| **first** | 0.004365 |
| **like** | 0.004067 |
| **hand** | 0.004067 |
| **made** | 0.003968 |
| **without** | 0.003869 |
| **see** | 0.003869 |
| **head** | 0.003869 |
| **come** | 0.003769 |
| **move** | 0.003769 |
| **much** | 0.003769 |
| **clerk** | 0.003670 |
| **thought** | 0.003670 |
| **chief** | 0.003670 |
| **thing** | 0.003571 |
| **make** | 0.003571 |
| **turn** | 0.003571 |
| **use** | 0.003373 |
| **samsa** | 0.003373 |
| **day** | 0.003373 |
| **let** | 0.003273 |
| **mr** | 0.003273 |
| **away** | 0.003174 |
| **went** | 0.003174 |
| **work** | 0.003174 |
| **famili** | 0.003174 |
| **bed** | 0.003174 |
| **well** | 0.003075 |
| **soon** | 0.003075 |

3. (Optional) Stop words dictionary



|  |  |
| --- | --- |
|  | **TF Scores** |
| **gregor** | 0.030958 |
| **room** | 0.013817 |
| **even** | 0.010596 |
| **father** | 0.010596 |
| **sister** | 0.010492 |
| **door** | 0.010077 |
| **mother** | 0.009350 |
| **back** | 0.008622 |
| **time** | 0.007688 |
| **way** | 0.006545 |
| **look** | 0.006337 |
| **open** | 0.005714 |
| **get** | 0.005402 |
| **said** | 0.005298 |
| **go** | 0.005090 |
| **littl** | 0.005090 |
| **still** | 0.004675 |
| **first** | 0.004571 |
| **want** | 0.004571 |
| **like** | 0.004259 |
| **hand** | 0.004259 |
| **made** | 0.004155 |
| **without** | 0.004052 |
| **see** | 0.004052 |
| **head** | 0.004052 |
| **much** | 0.003948 |
| **move** | 0.003948 |
| **come** | 0.003948 |
| **clerk** | 0.003844 |
| **chief** | 0.003844 |
| **thought** | 0.003844 |
| **make** | 0.003740 |
| **turn** | 0.003740 |
| **thing** | 0.003740 |
| **day** | 0.003532 |
| **use** | 0.003532 |
| **samsa** | 0.003532 |
| **let** | 0.003428 |
| **work** | 0.003324 |
| **bed** | 0.003324 |
| **famili** | 0.003324 |
| **went** | 0.003324 |
| **away** | 0.003324 |
| **well** | 0.003220 |
| **came** | 0.003220 |
| **soon** | 0.003220 |
| **seem** | 0.003220 |
| **left** | 0.003220 |
| **though** | 0.003117 |
| **quit** | 0.003117 |