

## Word list and Postings

The file postings store the invert file for all the distinct words. The postings are in the format **<term, [document frequency], document id, [term frequency], locations>** as below:

```
aarhus[1] 18[2]: 75 513
aaron[1] 17[1]: 423
aberrationcorrected[1] 39[1]: 497
ability[5] 11[1]: 101 14[1]: 393 17[1]: 843 19[1]: 759 38[1]: 415
able[14] 2[1]: 609 4[1]: 688 6[1]: 511 8[1]: 102 9[1]: 595 12[1]: 159 13[1]: 261 15[2]:
332 398 31[1]: 19 34[2]: 57 397 35[1]: 97 36[1]: 569 38[1]: 212 39[1]: 471
abnormalities[1] 16[1]: 528
aboard[4] 0[1]: 127 4[2]: 246 824 7[1]: 183 9[2]: 256 381
above[2] 18[1]: 134 28[1]: 398
absence[1] 30[1]: 452
absolutely[1] 34[1]: 17
absorbed[1] 34[1]: 236
absorbs[1] 30[1]: 191
abstract[1] 26[1]: 118
abundance[1] 6[1]: 332
abundant[1] 19[1]: 166
abundantly[1] 19[1]: 627
academia[1] 19[1]: 145
academic[1] 37[2]: 494 551
academics[1] 18[1]: 28
academy[2] 9[2]: 53 174 17[1]: 58
accelerate[3] 0[1]: 267 35[1]: 362 36[1]: 667
accelerated[1] 36[1]: 95
accelerometer[1] 4[2]: 169 372
accept[2] 23[1]: 320 36[1]: 403
```

For example: *ability[5] 11[1]: 101 14[1]: 393 17[1]: 843 19[1]: 759 38[1]: 415*

It means, for term “ability”, there are 5 documents containing the term, document 11,14,17,19 and 38. It occurs once in document11 at location 101, once in document14 at location 393, once in document17 at location 843, once in document19 at location 759 and once in document38 at location 415.

I have written the whole word list and postings to files by the word’s initial letter, and the files are stored in “postings” directory of the project. So 26 files from “a.txt” to “z.txt” to store all the postings.