

COMP 479 - Project 2

François LaBerge

October 23, 2020

1 Running the Program

See the “README.md” file packaged with the application.

2 Table Recreation

| | (distinct) terms | | | nonpositional postings | | | tokens | | |
|----------------|------------------|------------|-----|------------------------|------------|-----|-----------|------------|-----|
| | number | $\Delta\%$ | T% | number | $\Delta\%$ | T% | number | $\Delta\%$ | T% |
| unfiltered | 60,972 | | | 1,783,275 | | | 2,871,145 | | |
| no numbers | 58,759 | -4 | -4 | 1,621,515 | -9 | -9 | 2,644,815 | -8 | -8 |
| case folding | 42,720 | -27 | -30 | 1,498,356 | -8 | -16 | 2,644,815 | -0 | -8 |
| 30 stop words | 42,691 | -0 | -30 | 1,292,534 | -14 | -28 | 1,955,144 | -26 | -32 |
| 150 stop words | 42,571 | -0 | -30 | 1,154,273 | -23 | -35 | 1,756,076 | -34 | -39 |
| stemming | 31,844 | -25 | -48 | 1,096,885 | -5 | -38 | 1,756,076 | -0 | -39 |

3 Code Design