

Practical Exercise #1 grading criteria

Note that this breakdown is slightly different from what was listed on the assignment handout. Also note that although the assignment listed the analyzer component as optional, every group did it, so it was simpler to include it as a standard part of the grade.

Design (60 points)

Similarity (40 points)

- Tried a range of reasonable alternatives for the following and discovered optimal settings:
 - o tf
 - o idf
 - o coord
 - o lengthNorm
 - o boost
- minus: achieved no effect for tf/idf/coord

Analyzer (20 points):

- Porter stemming
- stopwords (disabling, default, english.stop, other)

Precision (15 points)

>42: 15

41-42: 13

40-41: 11

39-40: 8

38-39: 3

<38: 0

Coding style (5 points)

Writeup (20 points)

- justification for choices (10 points)
- test results (10 points)

Bonuses

- test script for finding optimal settings (up to 5 points)
- custom stopwords based on corpus (up to 5 points)
- extra filters (up to 5 points)