Composed by yu zhang(001259692)

Discussion

After comparing the result from both Lucene and BM25, we can find that even if the top 5 documents differ a little between both retrieval module in docID(different documents on top 5) and ranking order(same documents on top 5 in different order), their results are quite similar.

From the observation described above, we can get the same assumption:

1. As Lucene is widely used in both academic and commercial SE applications while our BM25 is only a retrieval model based SE, Lucene wins BM25 a lot. But BM25 is good enough (at least for what we discussed in class)
2. Our BM25 SE only based on retrieval model, not to say the information relevance is not considered in this implementation. But Lucene might have taken combing evidence into consideration, where PageRank, info relevance and other model like topic model are also important factor of great influence on document ranking.

top 5 rank

lucene_global warming potential

1. Globalwarming.txt score=0.33567125
2. UnitedNationsFrameworkConventiononClimateChange.txt score=0.20961201
3. Environmentalimpactoftheenergyindustry.txt score=0.19954512
4. Sustainabilityandenvironmentalmanagement.txt score=0.1823425
5. Climatechangemitigation.txt score=0.17927101

BM25_global warming potential:

1. Environmentalimpactoftheenergyindustry.txt Score=70.01280461119109
2. Climatechangemitigation.txt Score=68.13808104951019
3. Sustainabilityandenvironmentalmanagement.txt Score=67.96473966425334
4. Naturalenvironment.txt Score=65.58951666821244
5. UnitedNationsFrameworkConventiononClimateChange.txt Score=65.16073619766671

lucene_green power renewable energy

1. RenewableenergyintheCzechRepublic.txt score=0.5160874
2. 3Degrees.txt score=0.43727934
3. RenewableenergyinMexico.txt score=0.40383142
4. RenewableenergyinLuxembourg.txt score=0.36866862
5. RenewableenergyinIndia.txt score=0.3396703

BM25_green power renewable energy

1. Sustainableenergy.txt Score=104.18381826508778
2. CommunityChoiceAggregation.txt Score=99.81469770971802
3. SiliconValleyPower.txt Score=99.0641805398792
4. RenewableenergyinCanada.txt Score=98.9679436479127
5. Renewableenergycommercialization.txt Score=98.36481450109275

lucene_solar energy California
1. NevadaSolarOne.txt score=0.37490258
2. SolarDecathlon.txt score=0.35391805
3. Solarenergy.txt score=0.3228442
4. RenewableenergyintheUnitedStates.txt score=0.3134168
5. SiliconValleyPower.txt score=0.30913532

BM25_solar energy California
1. RenewableenergyintheUnitedStates.txt Score=76.4188851520471
2. SiliconValleyPower.txt Score=74.23793261133834
3. Netmetering.txt Score=74.11205440002183
4. Solarpower.txt Score=74.003262800594
5. SolarDecathlon.txt Score=73.84274232050824

lucene_light bulb bulbs alternative alternatives
1. Phaseoutofincandescentlightbulbs.txt score=0.6381232
2. Incandescentlightbulb.txt score=0.38270846
3. Compactfluorescentlamp.txt score=0.24038105
4. Energyconservation.txt score=0.2288304
5. Energysavinglamp.txt score=0.20609711

BM25_light bulb bulbs alternative alternatives
1. Phaseoutofincandescentlightbulbs.txt Score=105.49575478257981
2. Energyconservation.txt Score=81.62368710215227
3. Incandescentlightbulb.txt Score=79.52013506905503
4. Greenmarketing.txt Score=79.17348725643173
5. Lighting.txt Score=70.98959793470492