

COMS 435 PA3 Report

Brad Warren 184030907

PageRank:

- $\beta = 0.85$, approximation parameter = 0.01. Getting the top 10 documents is done in 7 iterations.
- $\beta = 0.85$, approximation parameter = 0.001. Getting the top 10 documents is done in 14 iterations.
- $\beta = 0.25$, approximation parameter = 0.01. Getting the top 10 documents is done in 3 iterations.
- $\beta = 0.25$, approximation parameter = 0.001. Getting the top 10 documents is done in 4 iterations.

Top 10 pages and page ranks:

- $\beta = 0.85$, approximation parameter = 0.01.
 - /wiki/Wayback_Machine 0.004632259449686032
 - /wiki/United_States 0.004127433115797951
 - /wiki/Association_football 0.0029708943164693333
 - /wiki/World_War_II 0.0026583541709091557
 - /wiki/France 0.002281316685790101
 - /wiki/National_Diet_Library 0.0021932861251187916
 - /wiki/Basketball 0.002176755573533508
 - /wiki/Canada 0.0021542403058700654
 - /wiki/The_New_York_Times 0.002054421597166199
 - /wiki/Japan 0.0020481994262459007
- $\beta = 0.85$, approximation parameter = 0.001.
 - /wiki/Wayback_Machine 0.004736316209802988
 - /wiki/United_States 0.004213755204811264
 - /wiki/Association_football 0.0029898910096220997
 - /wiki/World_War_II 0.0027537397053030337
 - /wiki/France 0.002351368070678008
 - /wiki/National_Diet_Library 0.002279046770532165
 - /wiki/Basketball 0.0021999017311340194
 - /wiki/Canada 0.002195436197074729
 - /wiki/Japan 0.002129454897323384
 - /wiki/The_New_York_Times 0.0021035440998979707
- $\beta = 0.25$, approximation parameter = 0.01.
 - /wiki/Wayback_Machine 0.0014515698320980313
 - /wiki/Association_football 0.0011646047781795495
 - /wiki/United_States 0.0011521216343682611
 - /wiki/Basketball 8.369450077587252E-4
 - /wiki/American_football 7.9099950724849E-4

- /wiki/National_Football_League 7.537317737530151E-4
- /wiki/Ice_hockey 7.50442982502212E-4
- /wiki/Baseball 6.661228912546342E-4
- /wiki/The_New_York_Times 6.060726195151414E-4
- /wiki/Canada 6.039393085451009E-4
- $\beta = 0.25$, approximation parameter = 0.001.
 - /wiki/Wayback_Machine 0.001452603079476902
 - /wiki/Association_football 0.0011644638620452276
 - /wiki/United_States 0.0011528928042926828
 - /wiki/Basketball 8.369783359897039E-4
 - /wiki/American_football 7.907422308271356E-4
 - /wiki/National_Football_League 7.535869353976607E-4
 - /wiki/Ice_hockey 7.5033868269537E-4
 - /wiki/Baseball 6.663356135400944E-4
 - /wiki/The_New_York_Times 6.066738777443723E-4
 - /wiki/Canada 6.043620319164678E-4

Index:

Files listed most relevant to least

Document name, TPScore, VSScore, RelevanceScore

- Query = "season" =>
 - 1944_Major_League_Baseball_season.txt, 0.058823529411764705, 0.09248859599841287, 0.07228955604642398
 - 1909_Major_League_Baseball_season.txt, 0.058823529411764705, 0.08734585770305087, 0.07023246072827917
 - 1908_Major_League_Baseball_season.txt, 0.058823529411764705, 0.08579064004268264, 0.06961037366413188
 - 1918_Major_League_Baseball_season.txt, 0.058823529411764705, 0.08567719822881349, 0.06956499693858423
 - 1997_Major_League_Baseball_season.txt, 0.058823529411764705, 0.08496183939059579, 0.06927885340329715
 - 1934_Major_League_Baseball_season.txt, 0.058823529411764705, 0.0785385492937141, 0.06670953736454446
 - 1940_Major_League_Baseball_season.txt, 0.058823529411764705, 0.07716872139751808, 0.06616160620606606
 - 1958_Major_League_Baseball_season.txt, 0.058823529411764705, 0.07670468326798399, 0.06597599095425243
 - 1924_Major_League_Baseball_season.txt, 0.058823529411764705, 0.07576338597599074, 0.06559947203745511
 - 1954_Major_League_Baseball_season.txt, 0.058823529411764705, 0.074433603900238, 0.06506755920715403

- Query = “baseball league” =>
 - Texas%E2%80%93Louisiana_League_(Negro_league).txt, 2.0, 0.041573702487105665, 1.2166294809948421
 - Miyazaki_Phoenix_League.txt, 2.0, 0.041323306556041925, 1.2165293226224168
 - Middle_States_League.txt, 2.0, 0.033242569978326884, 1.2132970279913307
 - Italian_Baseball_League_2D.txt, 2.0, 0.03315005954071471, 1.2132600238162858
 - Professional_baseball.txt, 2.0, 0.03289845820029599, 1.2131593832801184
 - Fieras_del_San_Fernando.txt, 2.0, 0.03223350094582144, 1.2128934003783285
 - Luis_Mel%C3%A9ndez_(baseball).txt, 2.0, 0.031536082590335114, 1.212614433036134
 - Liga_Paralela_de_B%C3%A9isbol_en_Venezuela.txt, 2.0, 0.030852917698759977, 1.212341167079504
 - Japan_Women%27s_Baseball_League.txt, 2.0, 0.030299595629176317, 1.2121198382516705
 - New_York_Collegiate_Baseball_League.txt, 2.0, 0.029635300044522177, 1.2118541200178088
- Query = “participated university the” =>
 - Baseball_at_the_Australian_University_Games.txt, 3.0, 0.12559936086891343, 1.850239744347565
 - Fordham_Rams.txt, 3.0, 0.04811752381652172, 1.8192470095266084
 - Yeshiva_University.txt, 3.0, 0.044644337965679515, 1.8178577351862717
 - University_of_Missouri.txt, 3.0, 0.041021880115786585, 1.8164087520463144
 - University_of_Michigan.txt, 3.0, 0.03987412653991099, 1.8159496506159643
 - University_of_Maryland,_College_Park.txt, 3.0, 0.039063791324082685, 1.8156255165296329
 - Sports_in_New_Jersey.txt, 3.0, 0.038482767062764316, 1.8153931068251055
 - Duquesne_University.txt, 3.0, 0.03842490463101232, 1.8153699618524048
 - Syracuse_Orange.txt, 3.0, 0.03835017069470308, 1.815340068277881
 - Buffalo_Bulls.txt, 3.0, 0.03789749028203415, 1.8151589961128134
- Query = “pitcher team professional residents developments”
 - Professional_sports_in_Canada.txt, 5.0, 0.02234293933196299, 3.008937175732785
 - History_of_baseball_outside_the_United_States.txt, 5.0, 0.02208342282208261, 3.008833369128833
 - London_Tigers.txt, 5.0, 0.0193511217014107, 3.0077404486805643
 - Toronto_sports.txt, 5.0, 0.01907633555200166, 3.0076305342208007
 - Cardiff.txt, 5.0, 0.016134536137356117, 3.0064538144549426
 - Brad_Thomas.txt, 2.5, 0.03970258058146325, 1.5158810322325853
 - Designated_hitter.txt, 2.5, 0.035313726985019285, 1.5141254907940078
 - Hold_(baseball).txt, 2.5, 0.03083230614110985, 1.512332922456444

- Adam_Loewen.txt, 2.5, 0.030725067854908, 1.5122900271419633
- Cuban_Stars_(East).txt, 2.5, 0.03061918396363399, 1.5122476735854535
- Query = “announcer Yankee broadcaster contract premiered”
 - Mel_Allen.txt, 0.29411764705882354, 0.05370030260122407, 3.0363813937774586
 - Bobby_Murcer.txt, 0.29411764705882354, 0.026519726844409126, 3.0236195131136436
 - Yankee_Stadium_(1923).txt, 0.29411764705882354, 0.008420496535599575, 1.5230211520176031
 - Ken_Singleton.txt, 0.29411764705882354, 0.029695214052812596, 1.521532653225475
 - Tampa_Bay_Rays.txt, 0.29411764705882354, 0.023660357623072878, 1.5139878316859197
 - Billy_Crystal.txt, 0.29411764705882354, 0.023457293578399447, 0.6161203626621982
 - ESPN2.txt, 0.8333333333333334, 0.02987425007117033, 0.5119497000284682
 - Paul_Olden.txt, 0.29411764705882354, 0.08656187063784973, 0.4757821846908787
 - Major_League_Baseball_on_ESPN_Radio.txt, 0.7142857142857143, 0.0630292955900526, 0.4537831468074496
 - Major_League_Baseball_on_regional_sports_networks.txt, 0.29411764705882354, 0.0380196052913559, 0.44817273263636837

I do think the ranking produced are acceptable because TPScore is at a whole number when we have a distance of 1 between letter and I looked through some documents that got whole numbers and it lined up. For VSScore the numbers returned all seem to be correct and for relevance as long as the TPScore and VSScore work well so will the relevance method. All and all, I believe the class is well implemented.