Project 4

Collin Vincent

Ryan McCullough

I think that all of our results made sense. The query with one word consistently had 0 for the term proximity because there was only one word so it can't have a proximity. With the exception of the last query the rest of the queries had results with identical Term Proximity which was presumably the highest term proximity among all documents and then are listed based on the the highest Vector Score which makes sense.

The last query started with the highest Term Proximity results then they slowly started to drop. It still followed the same pattern where the higher the Term Proximity the higher it was in the list until the last record which had a higher term proximity than the two records above it. This was because Albert_Pujols.txt had the phrase "all time" occur twice in the document and the words mlb and leader scattered lightly throughout the document. This document was very large though and the sheer amount of text caused the Vector Score to suffer, while the other two documents had far less text making it easier for the Vector Score to be large with only a handful of occurrences of each search term.

Results are listed below.

Query "valley"

Illinois_State_Redbirds.txt	VS= 0.1365691120856963 TP= 0.0 Relevance= 0.054627644834278524
Shenandoah_Valley.txt	VS= 0.13490099682252088 TP= 0.0 Relevance= 0.05396039872900835
Lehigh_Valley.txt	VS= 0.1313850670105557 TP= 0.0 Relevance= 0.052554026804222276
1996_Texas_Rangers_season.txt	VS= 0.12450350847336238 TP= 0.0 Relevance= 0.04980140038934495
Valley_Stream,_New_York.txt	VS= 0.12299961047348679 TP= 0.0 Relevance= 0.049199844189394715
George_Bird_(baseball).txt	VS= 0.12283929319642799 TP= 0.0 Relevance= 0.049135717278571196
Happy_Valley,_Pennsylvania.txt	VS= 0.11481333677192691 TP= 0.0 Relevance= 0.04592533470877077
Hudson_Valley.txt	VS= 0.1144135917221769 TP= 0.0 Relevance= 0.04576543668887076
Palm_Desert_Coyotes.txt	VS= 0.11194595035039055 TP= 0.0 Relevance= 0.044778380140156226
1955_Chicago_Cubs_season.txt	VS= 0.1074595353220668 TP= 0.0

	Relevance= 0.04298381412882672

Query "successful team"

industriales.txt	VS= 0.05424084610160469 TP= 2.0 Relevance= 1.2216963384406418
Jacksonville_Braves.txt	VS= 0.05005233226586082 TP= 2.0 Relevance= 1.2200209329063443
Philadelphia_Phillies_(NFL).txt	VS= 0.048181441757368264 TP= 2.0 Relevance= 1.2192725767029473
Kevin_Malone_(baseball).txt	VS= 0.04554048523066949 TP= 2.0 Relevance= 1.2182161940922678
Cuba_national_baseball_team.txt	VS= 0.04277592869867298 TP= 2.0 Relevance= 1.217110371479469
Cuban_national_baseball_team.txt	VS= 0.04277592869867298 TP= 2.0 Relevance= 1.217110371479469
Leones_del_Escogido.txt	VS= 0.04177121368180266 TP= 2.0 Relevance= 1.216708485472721
Bradley_Braves.txt	VS= 0.034929150477484294

	TP= 2.0 Relevance= 1.2139716601909938
Tigres_del_Licey.txt	VS= 0.03420010222830467 TP= 2.0 Relevance= 1.2136800408913218
Sport_in_Western_Australia.txt	VS= 0.031285369404490254 TP= 2.0 Relevance= 1.212514147761796

Query "career batting average"

Alex_Arias.txt	VS= 0.10559246327457182 TP= 1.5 Relevance= 0.9422369853098287
List_of_Detroit_Tigers_team_records.txt	VS= 0.10000135231148069 TP= 1.5 Relevance= 0.9400005409245922
Manuel_Cueto.txt	VS= 0.09615060280013714 TP= 1.5 Relevance= 0.9384602411200548
Batting_average.txt	VS= 0.08032979389674888 TP= 1.5 Relevance= 0.9321319175586995
Spud_Davis.txt	VS= 0.07584822263813559 TP= 1.5 Relevance= 0.9303392890552542
Ike_Boone.txt	VS= 0.07387122096916125 TP= 1.5 Relevance= 0.9295484883876645
Alejandro_Oms.txt	VS= 0.06825693610108352 TP= 1.5 Relevance= 0.9273027744404333
Clyde_Barnhart.txt	VS= 0.06739082023214178 TP= 1.5 Relevance= 0.9269563280928567
George_McQuinn.txt	VS= 0.06519515275725614 TP= 1.5 Relevance= 0.9260780611029024
Jimmy_Ripple.txt	VS= 0.06482217535026215 TP= 1.5 Relevance= 0.9259288701401047

Query "winners of the game"

2013_American_League_Wild_Card_Game.txt	VS= 0.07737794919358917 TP= 1.3333333333333333 Relevance= 0.8309511796774356
Major_League_Baseball_postseason.txt	VS= 0.05718531252209799 TP= 1.3333333333333333 Relevance= 0.8228741250088392
MLB_postseason.txt	VS= 0.05718531252209799 TP= 1.3333333333333333 Relevance= 0.8228741250088392
Grand_Forks_International.txt	VS= 0.05144301813446175 TP= 1.3333333333333333 Relevance= 0.8205772072537847
Playoff_format.txt	VS= 0.04880411251290529 TP= 1.3333333333333333 Relevance= 0.819521645005162
National_Football_League_Kickoff_game.txt	VS= 0.047289679424100076 TP= 1.3333333333333333 Relevance= 0.81891587176964
List_of_Major_League_Baseball_awards.txt	VS= 0.043247897220378881 TP= 1.3333333333333333 Relevance= 0.8172991588881514
2011_Major_League_Baseball_All-Star_Game.txt	VS= 0.04286786281518353 TP= 1.3333333333333333 Relevance= 0.8171471451260733
2012_Major_League_Baseball_All-Star_Game.txt	VS= 0.04168696582791491 TP= 1.3333333333333333 Relevance= 0.8166747863311659
Long_Island_Ducks.txt	VS= 0.04085910720545482 TP= 1.3333333333333333 Relevance= 0.8163436428821819

Query: "all time leader mlb"

Mike_Stanton_(left-handed_pitcher).txt	VS= 0.03721090124402407 TP= 0.2 Relevance= 0.13488436049760963
Cincinnati_Reds_Hall_of_Fame_and_Museum.txt	VS= 0.03509118679381658 TP= 0.2 Relevance= 0.13403647471752664
History_of_the_Cincinnati_Reds.txt	VS= 0.026208994511405317 TP= 0.2 Relevance= 0.13048359780456212
1921_Detroit_Tigers_season.txt	VS= 0.0243619821653965 TP= 0.2 Relevance= 0.1297447928661586
Fordham_Rams.txt	VS= 0.020577632514011263 TP= 0.2 Relevance= 0.1282310530056045
History_of_the_Arizona_Diamondbacks.txt	VS= 0.03035582437199393 TP= 0.19047619047619047 Relevance= 0.12642804403451186
Jim_Rice.txt	VS= 0.024163865468324865 TP= 0.18181818181818182 Relevance= 0.11875645527823903
List_of_Major_League_Baseball_career_runs_scored_leaders.txt	VS= 0.14061677211390927 TP= 0.0784313725490196 Relevance= 0.10330553237497547
List_of_Major_League_Baseball_players_with_a_home_run_in_their_first_major_leagu	VS= 0.13504771091968104 TP= 0.0784313725490196

e_at_bat.txt	Relevance= 0.10107790789728419
Albert_Pujols.txt	VS= 0.028385656497799647 TP= 0.13793103448275862 Relevance= 0.09411288328877503