Project 2 Report

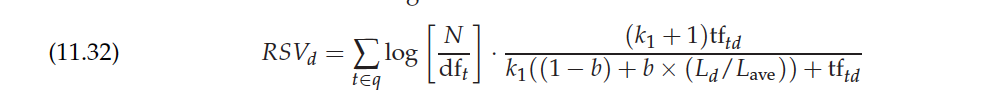
Shunyu Wang

40043915

COMP 479 Fall 2019

**Summary of Approach**

Given the following formula to calculate score of relevance of a document, my goal is to find all the missing variables based on project one.



As I did in project one, I create a pair of [id, cleaned\_token\_list] for each document.

* N – the total number of documents. I get this value by counting documents as processing each document. And persist this information in the “splitting\_word.txt”
* df – document frequency of a term is the length of its posting list in index file
* Lave – average length of document is the average length of cleaned token lists. This value is calculated by the sum of all lengths of token list diving N
* Ld – length of document could be found by finding length of the clean token list to which the term belongs
* tf – term frequency of a term is the occurrence in the given document. The value could be found during processing each document when constructing in-memory inverted-index

Since Ld and tf are more relevant to a single term, so we append them to id in the posting. So the posting list would contains many “id: Ld:tf” in block files.

**Additional Testing and Observations**

K is the smooth variable. As K grows from 0, to 0.5, to 1, to 1.5 to 3, the difference between adjacent rank score would increase. I could see the gap increase so the relevance is more distinguishable. A moderate K value should be selected for computation.

Additional Testing with old queries

**AND Jimmy Carter**

Document rank score: [(13540, 12.964350526063509), (18005, 11.35440658026538), (18438, 10.978583877097682), (18962, 10.287754619525208), (5525, 9.725401473120293), (10252, 9.670982759841444), (15158, 9.617169660223883), (21308, 9.53380569941816), (15473, 9.378604211288097), (5465, 8.981736772503078), (17363, 8.263934038635004), (1993, 7.876562130573988), (11330, 7.3323632924616025), (20614, 7.237380148045162), (19432, 6.362305734696884), (12136, 6.3532683685964635), (17023, 6.229388483972018), (11114, 6.1663605670548645), (16887, 6.1475606872772595), (12677, 6.001190194723038), (3380, 5.5243833025395634), (2534, 4.894174527735293), (11118, 4.412253622378426), (854, 2.8122732293262778), (965, 2.8122732293262778), (17325, 2.5653942817136652)]  
[INFO]----Results------ [13540, 18005, 18438, 18962, 5525, 10252, 15158, 21308, 15473, 5465, 17363, 1993, 11330, 20614, 19432, 12136, 17023, 11114, 16887, 12677, 3380, 2534, 11118, 854, 965, 17325]  
[INFO]----Total Count------ 26

Comparing this with the old AND Boolean result,

[INFO]----Results------ ['12136', '13540', '17023', '18005', '19432', '20614']

The new result provides more relevancy because it consider the weight of each term in query, but some highly ranked document may not contain both terms.

**OR Jimmy Carter**

In project one, we return the Ids based on their total counts in all postings. The total counts serve as Boolean relevancy score. However, in each rank level, we can’t find which term has better relevancy. By using this formula, each document has an relavency sore so it is possible to rank all documents.

**Results for Sample Queries**

**Test under K = 1 and b = 0.5**

1. Democrats’ welfare and healthcare reform policies

[INFO]----Total Count------ 612

[INFO] Query: ['democrats', 'healthcare', 'policies', 'reform', 'welfare', '’']  
[INFO] Mode: -r  
Document rank score: [(2132, 11.167684023893425), (6940, 10.948050825916614), (11204, 10.87265259280225), (18878, 10.506864289763236), (17560, 10.37598579121925), (8072, 10.326208666255884), (15630, 10.235532514857569), (1595, 10.218242763987876), (18731, 10.170045049370712), (1999, 10.103347740927813), (6276, 10.081999527134705), (5868, 9.997746025089683), (18271, 9.944226326882873), (21577, 9.8169828007464), (17697, 9.717589905672003), (7219, 9.673779629098611), (4006, 9.610112547702972), (19173, 9.579096937895207), (2916, 9.556397656051853), (19660, 9.488618693293372), (5093, 9.356844108709703), (17940, 9.325610016672048), (7375, 9.261743104047277), (14976, 9.089102543837432), (20805, 8.61217556659883), (12750, 8.430407301968918), (18357, 8.39008507920803), (20686, 8.384199190964411), (7467, 8.367926901190382), (8310, 8.366104955686264), (10230, 8.361480797605799), (219, 8.305167449431453), (15791, 8.263934038635004), (12806, 8.238779264459062), (20423, 8.230203695620167), (19133, 8.196747583036345), (9248, 8.186288728434231), (225, 8.17110080871543), (16092, 8.16510526675495), (9096, 8.06634310603959), (12326, 8.050243085070875), (20449, 8.033465758673072), (19453, 8.001586926297385), (7806, 7.84740231442342), (1760, 7.845914028898603), (15661, 7.835063631539106), (12552, 7.820423902908095), (6052, 7.8035306397045865), (18328, 7.766948786853057), (9347, 7.725737162261817), (13257, 7.725737162261817), (7433, 7.696182845140997), (10251, 7.666919792497874), (13671, 7.596189740333647), (2417, 7.593225416181234), (5659, 7.580450772281733), (6329, 7.495910429096435), (18722, 7.495910429096435), (3266, 7.4405900569259815), (1895, 7.424386918368178), (19134, 7.42407129160553), (19829, 7.3591237424340905), (17915, 7.338695971537851), (9688, 7.317328149278814), (8289, 7.305796758793264), (4268, 7.224267789826173), (3551, 7.221884738636047), (672, 7.04840895435661), (1908, 7.048232630352013), (18161, 7.020506390625947), (6600, 7.01091741736535), (19929, 7.001388560510212), (2594, 6.946724504570373), (16226, 6.941711751345308), (18449, 6.920972874526211), (18683, 6.900195293205946), (3289, 6.835528764964159), (951, 6.82452017034925), (862, 6.781344844582201), (20890, 6.765920529068393), (16171, 6.716265312373147), (19877, 6.710778167510097), (18597, 6.697715685025366), (3899, 6.6547851664255475), (2011, 6.59140273764401), (3619, 6.568077867840198), (1900, 6.552400354581029), (16544, 6.508901108467584), (2883, 6.427224160432205), (8139, 6.427224160432205), (8307, 6.427224160432205), (10212, 6.41718162268153), (2891, 6.412172113218078), (18941, 6.40148740682793), (9704, 6.38227853693268), (10125, 6.346143135299158), (16048, 6.3398224804820265), (11463, 6.309775552861914), (6606, 6.291419925845757), (5235, 6.236896337913711), (16778, 6.236896337913711), (18469, 6.236896337913711), (7401, 6.19107265190869), (16777, 6.185235298970149), (4130, 6.182574982712527), (7493, 6.166758625506079), (6079, 6.163389572861933), (20023, 6.121286003658511), (3113, 6.062166021482812), (7019, 6.044144574711081), (4396, 6.040121781404694), (4882, 6.024276716754256), (4287, 6.018237282196705), (15799, 6.018237282196705), (28, 5.9737223804291535), (12961, 5.9535250533558814), (12970, 5.9535250533558814), (12879, 5.953210830692763), (9055, 5.923831345908701), (19756, 5.911151209559754), (16168, 5.903343932671948), (2140, 5.869376289350851), (18159, 5.86623242017274), (13548, 5.848709471430601), (19147, 5.848709471430601), (3366, 5.833801475165027), (8242, 5.825692243250814), (11749, 5.807809405196821), (17185, 5.780931905546868), (2507, 5.747520726250141), (16827, 5.742173124277849), (20061, 5.720569764141364), (20091, 5.71606992138647), (11456, 5.708018865588628), (4059, 5.704161442383581), (2493, 5.688470855774967), (15666, 5.688470855774967), (17023, 5.688470855774967), (8592, 5.657019612281238), (8662, 5.657019612281238), (10664, 5.657019612281238), (2068, 5.649238292319366), (19115, 5.6115996279942255), (10355, 5.591657581432369), (9944, 5.573937885524464), (16196, 5.553747448973873), (11955, 5.536778299620969), (17934, 5.5335438406307995), (15873, 5.533214859202236), (1857, 5.5183836873298695), (1088, 5.509157403292661), (7892, 5.497968099375784), (7549, 5.432534707031983), (20461, 5.399600295840818), (19157, 5.38445917865117), (21455, 5.368622534008077), (21561, 5.361521309269303), (19039, 5.35811642337401), (7651, 5.316499985134212), (20795, 5.312814267326256), (19017, 5.3061966905893785), (20631, 5.235976044451703), (2802, 5.200162254817971), (19508, 5.200162254817971), (9774, 5.197310868024614), (20311, 5.1718620520706615), (2242, 5.166590021436542), (6581, 5.14661182771451), (16540, 5.140390274877534), (20462, 5.118650386270624), (8151, 5.117752322138541), (17242, 5.108204276761417), (19740, 5.108204276761417), (3502, 5.101858681386557), (3670, 5.101858681386557), (21158, 5.0986917920747095), (19600, 5.089214669784608), (15140, 5.079772713068867), (19220, 5.070365726563184), (1563, 5.0516558899338), (19321, 5.0516558899338), (1971, 5.02819921729092), (7031, 5.02819921729092), (12456, 5.02819921729092), (12471, 5.02819921729092), (61, 5.005479876314661), (18130, 4.998446559200442), (17402, 4.992623009335931), (19525, 4.98808855821709), (7595, 4.969142454345044), (9659, 4.960140384681376), (19102, 4.958308925033704), (7551, 4.933328815034448), (2156, 4.9288827593065), (20828, 4.9288827593065), (3847, 4.926424313084582), (17452, 4.911264351772528), (226, 4.871881658975302), (2980, 4.871881658975302), (234, 4.856823069842955), (4891, 4.856823069842955), (5160, 4.856823069842955), (9011, 4.856823069842955), (5288, 4.84266323582009), (1954, 4.837451399901271), (54, 4.820418120324154), (20094, 4.803504372533542), (11998, 4.786708902699509), (14270, 4.78153284059363), (18163, 4.78153284059363), (16292, 4.77418917583979), (4854, 4.77308474105256), (2541, 4.770030474467107), (17350, 4.759408280476477), (8137, 4.749345173190147), (18516, 4.7457161151959255), (16375, 4.737019882948303), (19834, 4.737019882948303), (11592, 4.71256022778678), (19824, 4.70446304526481), (10633, 4.701692774609692), (17394, 4.684311285609287), (15369, 4.683202837802705), (12415, 4.67566864486093), (3978, 4.665824409792951), (14874, 4.665672475981189), (12828, 4.658478686607764), (2415, 4.656458320313128), (821, 4.64067371583749), (17070, 4.632873372459352), (12497, 4.626834233666547), (4815, 4.609423387784713), (20335, 4.609423387784713), (21119, 4.607658919117498), (3752, 4.5909645027438835), (1105, 4.57859112432127), (7314, 4.574390623672607), (8135, 4.574390623672607), (428, 4.563329153906865), (2965, 4.563329153906865), (11320, 4.563329153906865), (12718, 4.548168591933753), (17747, 4.5415992930369615), (12640, 4.540626023787927), (15694, 4.525379295561828), (18868, 4.525379295561828), (17217, 4.510704270748137), (17271, 4.510704270748137), (14825, 4.509274742908584), (14905, 4.509274742908584), (2848, 4.503285349250197), (6039, 4.493284406940822), (11062, 4.477407076880961), (17087, 4.477407076880961), (12689, 4.46164155900462), (2052, 4.452879981299615), (11314, 4.4459866763414455), (20678, 4.4459866763414455), (4587, 4.444801123935259), (8413, 4.441266015738746), (17438, 4.4343174199730955), (1318, 4.43044126838221), (7216, 4.430416654278511), (15543, 4.416124987651806), (5227, 4.399674315129555), (12844, 4.399674315129555), (12856, 4.399674315129555), (16749, 4.399674315129555), (5774, 4.3942684917626735), (21525, 4.392049229661394), (16935, 4.3895134044710575), (16957, 4.3895134044710575), (2979, 4.387816494141217), (15191, 4.384450528572013), (2078, 4.371649998224498), (3124, 4.369331733645904), (6415, 4.369331733645904), (17953, 4.3603104017801515), (20727, 4.354316847963272), (3411, 4.3468480369033), (18507, 4.346027765625586), (4508, 4.339140241432835), (6108, 4.329520054979425), (20911, 4.325608046578882), (17307, 4.305004086378183), (18734, 4.2952752635841085), (17202, 4.285590313722588), (5891, 4.2647936017821175), (6426, 4.261567946493201), (13260, 4.259060230477668), (4940, 4.2515489011554655), (8835, 4.2515489011554655), (17038, 4.2515489011554655), (17423, 4.202673972661069), (17598, 4.200352053339156), (6431, 4.199382656969508), (4625, 4.1865405081714355), (4727, 4.1865405081714355), (17528, 4.1737766651587185), (8240, 4.168112509070106), (17803, 4.168112509070106), (3030, 4.161090413896837), (20860, 4.142228154329902), (1416, 4.131924104194753), (9694, 4.127381767939126), (7537, 4.123490199313433), (7633, 4.123490199313433), (2223, 4.113981177783481), (8038, 4.111107345862042), (17420, 4.100667322839198), (18885, 4.098798641233712), (19014, 4.098798641233712), (16006, 4.094042658343495), (7524, 4.086563421409133), (7610, 4.086563421409133), (3492, 4.074296471888463), (6927, 4.071160514427183), (1226, 4.06231081950166), (14749, 4.06231081950166), (3511, 4.061237829350358), (13722, 4.061237829350358), (3064, 4.035370072475516), (5453, 4.026466900506057), (2387, 4.002920310444619), (9226, 4.002920310444619), (6746, 3.9846107004946916), (21508, 3.9838635274062613), (32, 3.9681123077451006), (55, 3.9681123077451006), (17443, 3.9681123077451006), (19591, 3.9681123077451006), (16789, 3.9566437750637564), (11082, 3.9454477194448714), (6351, 3.910821613448493), (15408, 3.9047956941673716), (15456, 3.9047956941673716), (1175, 3.8868288428138404), (12744, 3.8749425160774065), (3928, 3.8631286669430245), (19291, 3.856334496259323), (20756, 3.839715765931248), (4575, 3.8281154161852324), (9442, 3.8281154161852324), (14950, 3.8165849480641927), (19976, 3.8165849480641927), (12623, 3.7937311459799763), (11163, 3.7824065753949614), (13869, 3.7824065753949614), (19398, 3.7824065753949614), (15382, 3.7820730974596435), (6127, 3.77114941296819), (13178, 3.77114941296819), (17196, 3.77114941296819), (18681, 3.77114941296819), (11881, 3.7599590586270386), (17017, 3.7599590586270386), (12438, 3.750681496361807), (19845, 3.7488349194003314), (20925, 3.7488349194003314), (10134, 3.737776409313605), (20907, 3.737776409313605), (10800, 3.7267829492862115), (15718, 3.7267829492862115), (12598, 3.723312074758999), (17269, 3.7158539670302404), (18041, 3.7158539670302404), (20027, 3.7158539670302404), (20829, 3.7049888969512046), (3364, 3.694187180050472), (17216, 3.694187180050472), (17277, 3.694187180050472), (21391, 3.694187180050472), (12440, 3.6900235745877614), (12612, 3.6900235745877614), (4736, 3.6834482638293955), (20780, 3.6830010862223737), (7343, 3.6727716021951076), (7360, 3.6727716021951076), (11387, 3.6621566553679537), (1376, 3.6604244550055074), (6954, 3.6516028897905244), (10661, 3.641109778038253), (10774, 3.641109778038253), (13045, 3.641109778038253), (16103, 3.6312964089444546), (570, 3.6306767987315527), (860, 3.621689804687988), (14911, 3.621689804687988), (15763, 3.6203034364494626), (3198, 3.602628279400157), (15224, 3.5997335305605453), (18, 3.59317256213139), (16701, 3.589535985148192), (4491, 3.5744092589348817), (10780, 3.5659649233320225), (18568, 3.5651009014897386), (7877, 3.559287086396944), (8390, 3.546628876093419), (15154, 3.5416492321669306), (6744, 3.529543740271064), (10100, 3.529543740271064), (9549, 3.528347283948607), (1957, 3.492343740234846), (2427, 3.492343740234846), (19015, 3.492343740234846), (8608, 3.490650696301135), (8676, 3.490650696301135), (8982, 3.490650696301135), (17381, 3.4834573693945536), (18009, 3.4834573693945536), (17382, 3.481060996586022), (3537, 3.474616107035684), (16963, 3.4715238431707176), (17477, 3.4614380434186085), (12431, 3.448359562146246), (15364, 3.4432233770579135), (19019, 3.4310745518096732), (5432, 3.4224968654301486), (2190, 3.415380600424021), (18443, 3.4061994697777207), (18472, 3.4061994697777207), (10486, 3.4035729025483894), (9180, 3.3970192212706194), (5973, 3.388610757851633), (6099, 3.3802438177087892), (5717, 3.378949874019499), (9462, 3.378949874019499), (5516, 3.369963305205617), (8141, 3.361024410496848), (8176, 3.343288133924306), (10670, 3.3390213321269746), (18985, 3.3390213321269746), (2913, 3.3344900072560844), (874, 3.3257380649798214), (8635, 3.3257380649798214), (8667, 3.3257380649798214), (6796, 3.3170319443908687), (10995, 3.3170319443908687), (15525, 3.3083712865726165), (21273, 3.3083712865726165), (1210, 3.2997557363471666), (5944, 3.2997557363471666), (12986, 3.2997557363471666), (19495, 3.298792159450746), (742, 3.282970614321486), (9701, 3.2826585563660933), (9873, 3.275116617636506), (3246, 3.265737635972453), (17058, 3.265737635972453), (1786, 3.26332709211452), (12519, 3.2573424235406483), (7155, 3.2517784944704506), (21542, 3.244830224730632), (8885, 3.2406808254662716), (926, 3.232413782544163), (942, 3.232413782544163), (7032, 3.232413782544163), (10605, 3.2211735204048466), (18726, 3.2211735204048466), (12039, 3.2136120802161026), (13637, 3.2136120802161026), (6392, 3.207863804448891), (13241, 3.206086056609039), (13323, 3.206086056609039), (20863, 3.1997631382760408), (10936, 3.1836839265259096), (16214, 3.154374991179861), (7310, 3.1520054794958012), (11820, 3.1520054794958012), (14880, 3.1327202429566583), (20868, 3.124799513581534), (11770, 3.113283053457769), (17293, 3.098059175445751), (17300, 3.098059175445751), (6472, 3.097282231158506), (9445, 3.0902906234132272), (209, 3.071772612744999), (14419, 3.064637435214559), (7099, 3.0558007727054903), (16051, 3.0532679584513542), (15552, 3.045928372012769), (15738, 3.042271795335683), (2118, 3.04221943593791), (17296, 3.0386239874276066), (14809, 3.0313545520509857), (2278, 3.0241198156499096), (18265, 3.0241198156499096), (17126, 3.0169195303745524), (181, 3.0097534507299573), (14766, 2.97442770600308), (9795, 2.9605285111152146), (5285, 2.95042833226737), (8203, 2.95042833226737), (13592, 2.9440833251012033), (10344, 2.9365149542463778), (17119, 2.93311621008637), (17363, 2.93311621008637), (3562, 2.9314748312035537), (8117, 2.926342269647372), (11327, 2.91959954552376), (10905, 2.919215272207892), (5318, 2.909543519534586), (18776, 2.906579078921571), (17395, 2.906206887058055), (17397, 2.906206887058055), (8743, 2.900421072398432), (7254, 2.8995565280487687), (9639, 2.8995565280487687), (8427, 2.8863467033196173), (7105, 2.860284882070682), (13908, 2.860284882070682), (17201, 2.847429669117555), (5169, 2.846151239442951), (5850, 2.8410452976621343), (18835, 2.8346894916270964), (20865, 2.811085721092525), (17312, 2.8095481214131084), (238, 2.799588437979672), (16198, 2.7971439354466052), (8747, 2.7726612751801136), (19947, 2.7726612751801136), (18625, 2.760579962434231), (18420, 2.7600781172823785), (11600, 2.7545787016463303), (5271, 2.748603476696989), (308, 2.7108886062813062), (18148, 2.7074918862335724), (237, 2.705531118917114), (263, 2.7017189824249717), (12602, 2.684547039740068), (18403, 2.679058211687005), (17750, 2.6788714646031964), (6410, 2.668613540296412), (3885, 2.667592005804867), (9307, 2.658249992567106), (12773, 2.6576535838180657), (17409, 2.6530983452946892), (9673, 2.6428547223398837), (10350, 2.62260296201544), (19016, 2.6125930270459157), (1867, 2.6072144089656626), (9782, 2.601860785949306), (18005, 2.596529104010885), (8076, 2.5754191112953495), (9848, 2.5672851466557813), (18455, 2.564992313273911), (12942, 2.544389965376128), (236, 2.5392909874896032), (13752, 2.5241159417476333), (11175, 2.507323710937838), (179, 2.499223279600221), (1611, 2.4894031488355837), (17305, 2.484571227172522), (6657, 2.472402346843536), (5542, 2.455632175886264), (740, 2.450882403785903), (2352, 2.436742697610215), (5206, 2.4320656482865872), (4603, 2.4227652060369254), (9788, 2.418141608315481), (16784, 2.4135356242996417), (15957, 2.4089471535310114), (19000, 2.404376096313685), (16759, 2.3684228088921717), (17410, 2.320707331057348), (19604, 2.3124978820473983), (10255, 2.299648280866265), (8069, 2.293130228093902), (7628, 2.234755207684854), (8210, 2.219100179960266), (17207, 2.219100179960266), (15678, 2.215220634191105), (3019, 2.184665866822952), (19211, 2.1730138828127927), (11732, 2.162297274329884), (13657, 2.162297274329884), (9758, 2.1424080534774017), (13046, 2.1159176911469237), (7062, 2.090934327982363), (13244, 2.0772232832087085), (12673, 2.070434971825673), (6190, 2.063690883969563), (2197, 2.033878335083968), (3065, 2.030618914675179), (5759, 2.024131314308805), (20500, 2.024131314308805), (12025, 2.022154721081329), (4871, 1.9985902251692618), (11198, 1.9923053502473458), (17358, 1.9840561538725503), (14826, 1.9697823685928917), (6395, 1.9669522215115798), (6722, 1.949394158088172), (19285, 1.9404382890617333), (6896, 1.9363490044866476), (8980, 1.9227711726211107), (9470, 1.899709449411262), (335, 1.89874999469079), (17368, 1.8799795293135193), (16200, 1.866135791984259), (13034, 1.8552067390297395), (15470, 1.841724131914698), (11768, 1.8284360790148801), (18299, 1.817942902090835), (10689, 1.812741348722907), (10771, 1.812741348722907), (16937, 1.7896980264933782), (15442, 1.7871737697564345), (17450, 1.7821465580271616), (18086, 1.7746585472791485), (15549, 1.6872253032720532), (7405, 1.6203404127331358)]  
[INFO]----Results------ [2132, 6940, 11204, 18878, 17560, 8072, 15630, 1595, 18731, 1999, 6276, 5868, 18271, 21577, 17697, 7219, 4006, 19173, 2916, 19660, 5093, 17940, 7375, 14976, 20805, 12750, 18357, 20686, 7467, 8310, 10230, 219, 15791, 12806, 20423, 19133, 9248, 225, 16092, 9096, 12326, 20449, 19453, 7806, 1760, 15661, 12552, 6052, 18328, 9347, 13257, 7433, 10251, 13671, 2417, 5659, 6329, 18722, 3266, 1895, 19134, 19829, 17915, 9688, 8289, 4268, 3551, 672, 1908, 18161, 6600, 19929, 2594, 16226, 18449, 18683, 3289, 951, 862, 20890, 16171, 19877, 18597, 3899, 2011, 3619, 1900, 16544, 2883, 8139, 8307, 10212, 2891, 18941, 9704, 10125, 16048, 11463, 6606, 5235, 16778, 18469, 7401, 16777, 4130, 7493, 6079, 20023, 3113, 7019, 4396, 4882, 4287, 15799, 28, 12961, 12970, 12879, 9055, 19756, 16168, 2140, 18159, 13548, 19147, 3366, 8242, 11749, 17185, 2507, 16827, 20061, 20091, 11456, 4059, 2493, 15666, 17023, 8592, 8662, 10664, 2068, 19115, 10355, 9944, 16196, 11955, 17934, 15873, 1857, 1088, 7892, 7549, 20461, 19157, 21455, 21561, 19039, 7651, 20795, 19017, 20631, 2802, 19508, 9774, 20311, 2242, 6581, 16540, 20462, 8151, 17242, 19740, 3502, 3670, 21158, 19600, 15140, 19220, 1563, 19321, 1971, 7031, 12456, 12471, 61, 18130, 17402, 19525, 7595, 9659, 19102, 7551, 2156, 20828, 3847, 17452, 226, 2980, 234, 4891, 5160, 9011, 5288, 1954, 54, 20094, 11998, 14270, 18163, 16292, 4854, 2541, 17350, 8137, 18516, 16375, 19834, 11592, 19824, 10633, 17394, 15369, 12415, 3978, 14874, 12828, 2415, 821, 17070, 12497, 4815, 20335, 21119, 3752, 1105, 7314, 8135, 428, 2965, 11320, 12718, 17747, 12640, 15694, 18868, 17217, 17271, 14825, 14905, 2848, 6039, 11062, 17087, 12689, 2052, 11314, 20678, 4587, 8413, 17438, 1318, 7216, 15543, 5227, 12844, 12856, 16749, 5774, 21525, 16935, 16957, 2979, 15191, 2078, 3124, 6415, 17953, 20727, 3411, 18507, 4508, 6108, 20911, 17307, 18734, 17202, 5891, 6426, 13260, 4940, 8835, 17038, 17423, 17598, 6431, 4625, 4727, 17528, 8240, 17803, 3030, 20860, 1416, 9694, 7537, 7633, 2223, 8038, 17420, 18885, 19014, 16006, 7524, 7610, 3492, 6927, 1226, 14749, 3511, 13722, 3064, 5453, 2387, 9226, 6746, 21508, 32, 55, 17443, 19591, 16789, 11082, 6351, 15408, 15456, 1175, 12744, 3928, 19291, 20756, 4575, 9442, 14950, 19976, 12623, 11163, 13869, 19398, 15382, 6127, 13178, 17196, 18681, 11881, 17017, 12438, 19845, 20925, 10134, 20907, 10800, 15718, 12598, 17269, 18041, 20027, 20829, 3364, 17216, 17277, 21391, 12440, 12612, 4736, 20780, 7343, 7360, 11387, 1376, 6954, 10661, 10774, 13045, 16103, 570, 860, 14911, 15763, 3198, 15224, 18, 16701, 4491, 10780, 18568, 7877, 8390, 15154, 6744, 10100, 9549, 1957, 2427, 19015, 8608, 8676, 8982, 17381, 18009, 17382, 3537, 16963, 17477, 12431, 15364, 19019, 5432, 2190, 18443, 18472, 10486, 9180, 5973, 6099, 5717, 9462, 5516, 8141, 8176, 10670, 18985, 2913, 874, 8635, 8667, 6796, 10995, 15525, 21273, 1210, 5944, 12986, 19495, 742, 9701, 9873, 3246, 17058, 1786, 12519, 7155, 21542, 8885, 926, 942, 7032, 10605, 18726, 12039, 13637, 6392, 13241, 13323, 20863, 10936, 16214, 7310, 11820, 14880, 20868, 11770, 17293, 17300, 6472, 9445, 209, 14419, 7099, 16051, 15552, 15738, 2118, 17296, 14809, 2278, 18265, 17126, 181, 14766, 9795, 5285, 8203, 13592, 10344, 17119, 17363, 3562, 8117, 11327, 10905, 5318, 18776, 17395, 17397, 8743, 7254, 9639, 8427, 7105, 13908, 17201, 5169, 5850, 18835, 20865, 17312, 238, 16198, 8747, 19947, 18625, 18420, 11600, 5271, 308, 18148, 237, 263, 12602, 18403, 17750, 6410, 3885, 9307, 12773, 17409, 9673, 10350, 19016, 1867, 9782, 18005, 8076, 9848, 18455, 12942, 236, 13752, 11175, 179, 1611, 17305, 6657, 5542, 740, 2352, 5206, 4603, 9788, 16784, 15957, 19000, 16759, 17410, 19604, 10255, 8069, 7628, 8210, 17207, 15678, 3019, 19211, 11732, 13657, 9758, 13046, 7062, 13244, 12673, 6190, 2197, 3065, 5759, 20500, 12025, 4871, 11198, 17358, 14826, 6395, 6722, 19285, 6896, 8980, 9470, 335, 17368, 16200, 13034, 15470, 11768, 18299, 10689, 10771, 16937, 15442, 17450, 18086, 15549, 7405]

1. Drug company bankruptcies

[INFO]----Total Count------ 5500

Top Ranks with scores

[(4050, 14.561068183268457), (16771, 13.135299604625503), (8384, 11.007501530243216), (8209, 10.892687477006515), (9342, 10.26341598458943), (192, 10.142380123060978), (19640, 10.068385670119932), (17680, 9.881052110563736), (9542, 9.68988570652803), (730, 9.402920280162578), (7940, 9.395376721522208), (1998, 9.245728582124595), (3108, 9.245728582124595), (20953, 9.226623129085013), (21348, 9.226623129085013), (16413, 9.0953160970819), (2523, 9.000732654716222), (2356, 8.932476350741053), (21051, 8.914780911455775), (5319, 8.899814416240174), (4871, 8.885419613667178), (6544, 8.812002844046228), (3739, 8.79827698291843), (1391, 8.781279753880861),…]

Top Ranks:

[INFO]----Results------ [4050, 16771, 8384, 8209, 9342, 192, 19640, 17680, 9542, 730, 7940, 1998, 3108, 20953, 21348, 16413, 2523, 2356, 21051, 5319, 4871, 6544, 3739, 1391,…]

1. George Bush

[INFO]----Total Count------ 144

Document rank score: [(20891, 13.90911547794298), (8593, 12.7233331030515), (4008, 11.931489179337143), (16780, 11.401907742646255), (20719, 10.760088826022782), (3560, 10.093838855820717), (7525, 9.503342708032939), (20860, 8.975347185285935), (8500, 8.801334792159413), (16824, 8.086196065453844), (2796, 7.6988321324196995), (2711, 7.537489903869477), (21516, 7.227730044806347), (854, 6.439826183279685), (965, 6.439826183279685), (5405, 6.429343855265386), (4853, 5.851941163267275), (2766, 5.829433697254708), (5459, 5.762938255841156), (6564, 5.741108944266001), (10400, 5.741108944266001), (2733, 5.697942711602347), (5334, 5.676602102195597), (10682, 5.676602102195597), (16115, 5.676602102195597), (21393, 5.676602102195597), (7469, 5.655420751068001), (15284, 5.613528745504534), (16229, 5.551841616433054), (255, 5.511464586495361), (15478, 5.511464586495361), (4737, 5.432447173068905), (1667, 5.413045576022229), (43, 5.39378206863425), (6065, 5.39378206863425), (16318, 5.39378206863425), (12009, 5.355663467442488), (17150, 5.355663467442488), (18160, 5.355663467442488), (871, 5.336805497486704), (286, 5.299485179322462), (8053, 5.281020074168029), (8554, 5.281020074168029), (8748, 5.281020074168029), (2356, 5.2626831989105005), (13507, 5.2626831989105005), (6423, 5.226388832021463), (17647, 5.190591648240493), (11380, 5.155281500973552), (16730, 5.120448517858866), (12605, 4.936979678456756), (16575, 4.936979678456756), (9002, 4.920950523656573), (15880, 4.857861414378923), (5574, 4.842341090371323), (14914, 4.766203651843472), (15342, 4.721659692480449), (6989, 4.620892564897025), (15405, 4.57901136340249), (15427, 4.57901136340249), (15439, 4.57901136340249), (19003, 4.537882518821031), (4098, 4.510871313351857), (5157, 4.510871313351857), (19033, 4.49748593853479), (6296, 4.444729505238194), (1502, 4.418812715120187), (12460, 4.380499310075792), (17181, 4.380499310075792), (18867, 4.380499310075792), (16550, 4.367875392755689), (3014, 4.293633884663524), (6882, 4.245525941978219), (8080, 4.237946130689665), (8198, 4.237946130689665), (18330, 4.2336669309671064), (6744, 4.221873986869705), (9247, 4.2101465591284), (1729, 4.107460057686244), (16218, 4.096358814287092), (1022, 4.074335379801678), (18443, 4.074335379801678), (18472, 4.074335379801678), (6940, 4.063412228649394), (16090, 4.030991386399532), (16199, 4.030991386399532), (21376, 4.030991386399532), (15363, 3.9676773855660317), (17356, 3.947012399182875), (3938, 3.9367604189252576), (12261, 3.9265615577363318), (18442, 3.916415403840372), (9150, 3.8763497731105483), (18579, 3.8664611257301633), (9342, 3.856622802255023), (6062, 3.817765141778902), (8765, 3.798628474401564), (3366, 3.7516157457579804), (18688, 3.7239625584428113), (9755, 3.6787688380733594), (16389, 3.6787688380733594), (7186, 3.66099700793774), (925, 3.6433960607841924), (14809, 3.6259635437469475), (17004, 3.5915942210574032), (17862, 3.589280906604513), (20099, 3.5831034545773615), (14749, 3.520680049445352), (2802, 3.5165957338427476), (17119, 3.5084554659403335), (20571, 3.4136323452392436), (13246, 3.3907220610430073), (21379, 3.3681172473027203), (5711, 3.3384422054762646), (7765, 3.2948973071439656), (19828, 3.287750024482048), (18148, 3.2385742762526153), (7471, 3.2316690006102857), (12720, 3.211128731538611), (19958, 3.1058458223078365), (178, 3.093168900584131), (19485, 3.007247542234572), (19760, 2.966052370423139), (12709, 2.8869576405451887), (11624, 2.8015762685512464), (17305, 2.7507309642218223), (3563, 2.721100110029128), (8009, 2.7065227880111147), (8252, 2.696891034317125), (3390, 2.673108926430731), (7326, 2.6267812154007353), (342, 2.604214366471176), (10725, 2.604214366471176), (18961, 2.444601227881007), (12806, 2.4211705451856615), (12002, 2.38685474218303), (12209, 2.292969381673562), (28, 2.2757548968261627), (6896, 2.143780426147418), (6656, 2.1257401981573967), (4764, 2.0992420516429697), (19589, 2.042658707932916), (12967, 2.004831694823048), (5631, 1.9735061995914376)]

[INFO]----Results------ [20891, 8593, 4008, 16780, 20719, 3560, 7525, 20860, 8500, 16824, 2796, 2711, 21516, 854, 965, 5405, 4853, 2766, 5459, 6564, 10400, 2733, 5334, 10682, 16115, 21393, 7469, 15284, 16229, 255, 15478, 4737, 1667, 43, 6065, 16318, 12009, 17150, 18160, 871, 286, 8053, 8554, 8748, 2356, 13507, 6423, 17647, 11380, 16730, 12605, 16575, 9002, 15880, 5574, 14914, 15342, 6989, 15405, 15427, 15439, 19003, 4098, 5157, 19033, 6296, 1502, 12460, 17181, 18867, 16550, 3014, 6882, 8080, 8198, 18330, 6744, 9247, 1729, 16218, 1022, 18443, 18472, 6940, 16090, 16199, 21376, 15363, 17356, 3938, 12261, 18442, 9150, 18579, 9342, 6062, 8765, 3366, 18688, 9755, 16389, 7186, 925, 14809, 17004, 17862, 20099, 14749, 2802, 17119, 20571, 13246, 21379, 5711, 7765, 19828, 18148, 7471, 12720, 19958, 178, 19485, 19760, 12709, 11624, 17305, 3563, 8009, 8252, 3390, 7326, 342, 10725, 18961, 12806, 12002, 12209, 28, 6896, 6656, 4764, 19589, 12967, 5631]