Information Retrieval HW2 Part 2

1. Please produce a table containing a table of results

term	stem	removestop	sim	termweights		p_0.25	p_0.5	p_0.75	p_1.0	p_mean1 p_mean2
	r_norm	p_norm								
tf	False	False	cosine	1,1,1,1	0.1897	0.0488	0.0275	0.0061	0.0887	0.1064	0.7706 0.4013
tf	False	False	cosine	3,3,4,1	0.2050	0.0601	0.0312	0.0058	0.0988	0.1150	0.7592 0.4032
tf	False	False	cosine	1,1,1,4	0.1218	0.0369	0.0218	0.0061	0.0602	0.0783	0.7510 0.3603
tf	False	False	jaccard	1,1,1,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796 0.1004
tf	False	False	jaccard	3,3,4,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796 0.1004
tf	False	False	jaccard	1,1,1,4	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796 0.1004
tf	False	False	dice	1,1,1,1	0.1296	0.0401	0.0232	0.0064	0.0643	0.0793	0.7619 0.3649
tf	False	False	dice	3,3,4,1	0.1587	0.0580	0.0237	0.0061	0.0801	0.0939	0.7511 0.3673
tf	False	False	dice	1,1,1,4	0.1193	0.0356	0.0212	0.0060	0.0587	0.0752	0.7663 0.3578
tf	False	False	overlap	1,1,1,1	0.1270	0.0477	0.0235	0.0060	0.0661	0.0817	0.7604 0.3659
tf	False	False	overlap	3,3,4,1	0.1874	0.0757	0.0271	0.0065	0.0968	0.1091	0.7718 0.3979
tf	False	False	overlap	1,1,1,4	0.1069	0.0357	0.0220	0.0060	0.0548	0.0709	0.7472 0.3392
tf	False	True	cosine	1,1,1,1	0.2468	0.0877	0.0421	0.0079	0.1255	0.1367	0.8012 0.4674
tf	False	True	cosine	3,3,4,1	0.2385	0.0893	0.0461	0.0074	0.1247	0.1401	0.7840 0.4512
tf	False	True	cosine	1,1,1,4	0.1953	0.0659	0.0358	0.0082	0.0990	0.1121	0.7895 0.4392
tf	False	True	jaccard	1,1,1,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796 0.1004

tf	False	True	jaccard	3,3,4,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tf	False	True	jaccard	1,1,1,4	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tf	False	True	dice	1,1,1,1	0.1665	0.0719	0.0342	0.0093	0.0909	0.1042	0.7906
0.4281											
tf	False	True	dice	3,3,4,1	0.1755	0.0724	0.0370	0.0081	0.0949	0.1061	0.7700
0.4095											
tf	False	True	dice	1,1,1,4	0.1898	0.0737	0.0364	0.0095	0.1000	0.1109	0.8005
0.4445											
tf	False	True	overlap	1,1,1,1	0.1839	0.0856	0.0374	0.0080	0.1023	0.1130	0.7895
0.4341											
tf	False	True	overlap	3,3,4,1	0.2135	0.0989	0.0473	0.0090	0.1199	0.1283	0.7838
0.4377											
tf	False	True	overlap	1,1,1,4	0.1502	0.0583	0.0325	0.0076	0.0803	0.0939	0.7802
0.4098											
tf	True	False	cosine	1,1,1,1	0.2155	0.0948	0.0549	0.0088	0.1217	0.1334	0.8160
0.4589											
tf	True	False	cosine	3,3,4,1	0.2174	0.0858	0.0375	0.0080	0.1136	0.1295	0.8064
0.4457											
tf	True	False	cosine	1,1,1,4	0.1730	0.0707	0.0372	0.0090	0.0936	0.1084	0.7915
0.4233											
tf	True	False	jaccard	1,1,1,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tf	True	False	jaccard	3,3,4,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tf	True	False	jaccard	1,1,1,4	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tf	True	False	dice	1,1,1,1	0.1669	0.0750	0.0284	0.0078	0.0901	0.0991	0.8037
0.4076											
tf	True	False	dice	3,3,4,1	0.1694	0.0664	0.0247	0.0076	0.0868	0.0989	0.7976
0.4053											
tf	True	False	dice	1,1,1,4	0.1581	0.0530	0.0251	0.0080	0.0787	0.0927	0.8030
0.4028											
tf	True	False	overlap	1,1,1,1	0.1405	0.0643	0.0279	0.0082	0.0776	0.0934	0.7895
0.3955											

tf	True	False	overlap	3,3,4,1	0.1615	0.0598	0.0296	0.0088	0.0836	0.0991	0.8005
0.4208											
tf	True	False	overlap	1,1,1,4	0.1187	0.0474	0.0264	0.0081	0.0641	0.0809	0.7768
0.3695											
tf	True	True	cosine	1,1,1,1	0.2648	0.1178	0.0541	0.0115	0.1456	0.1609	0.8520
0.5129											
tf	True	True	cosine	3,3,4,1	0.2359	0.0966	0.0448	0.0098	0.1257	0.1496	0.8371
0.4953											
tf	True	True	cosine	1,1,1,4	0.2332	0.1095	0.0559	0.0188	0.1329	0.1474	0.8348
0.4879											
tf	True	True	jaccard	1,1,1,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tf	True	True	jaccard	3,3,4,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tf	True	True	jaccard	1,1,1,4	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tf	True	True	dice	1,1,1,1	0.2347	0.1295	0.0537	0.0099	0.1393	0.1458	0.8406
0.4759											
tf	True	True	dice	3,3,4,1	0.1985	0.0827	0.0312	0.0096	0.1041	0.1188	0.8271
0.4495											
tf	True	True	dice	1,1,1,4	0.2238	0.1222	0.0590	0.0110	0.1350	0.1415	0.8454
0.4852											
tf	True	True	overlap	1,1,1,1	0.1741	0.0807	0.0412	0.0106	0.0987	0.1145	0.8300
0.4642											
tf	True	True	overlap	3,3,4,1	0.2030	0.0820	0.0429	0.0101	0.1093	0.1219	0.8251
0.4656											
tf	True	True	overlap	1,1,1,4	0.1608	0.0727	0.0378	0.0102	0.0904	0.1053	0.8207
0.4412											
tfidf	False	False	cosine	1,1,1,1	0.5528	0.3411	0.1685	0.0664	0.3541	0.3339	0.8905
0.6718											
tfidf	False	False	cosine	3,3,4,1	0.5430	0.3423	0.1742	0.0541	0.3532	0.3325	0.8905
0.6681											
tfidf	False	False	cosine	1,1,1,4	0.4540	0.2506	0.1339	0.0434	0.2795	0.2711	0.8824
0.6309											
tfidf	False	False	jaccard	1,1,1,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											

tfidf	False	False	jaccard	3,3,4,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tfidf	False	False	jaccard	1,1,1,4	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tfidf	False	False	dice	1,1,1,1	0.4254	0.2818	0.1689	0.0614	0.2920	0.2869	0.8872
0.6545											
tfidf	False	False	dice	3,3,4,1	0.4102	0.2801	0.1392	0.0523	0.2765	0.2666	0.8871
0.6452											
tfidf	False	False	dice	1,1,1,4	0.4155	0.2447	0.1502	0.0685	0.2701	0.2671	0.8796
0.6297											
tfidf	False	False	overlap	1,1,1,1	0.4470	0.2961	0.1604	0.0433	0.3012	0.2875	0.8895
0.6451											
tfidf	False	False	overlap	3,3,4,1	0.5018	0.2930	0.1743	0.0562	0.3230	0.3058	0.8898
0.6570											
tfidf	False	False	overlap	1,1,1,4	0.3765	0.2435	0.1195	0.0195	0.2465	0.2426	0.8809
0.5998											
tfidf	False	True	cosine	1,1,1,1	0.5609	0.3433	0.1790	0.0843	0.3611	0.3409	0.8888
0.6795											
tfidf	False	True	cosine	3,3,4,1	0.5617	0.3587	0.1913	0.0681	0.3706	0.3476	0.8870
0.6743											
tfidf	False	True	cosine	1,1,1,4	0.4824	0.2738	0.1413	0.0461	0.2992	0.2891	0.8798
0.6456											
tfidf	False	True	jaccard	1,1,1,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tfidf	False	True	jaccard	3,3,4,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tfidf	False	True	jaccard	1,1,1,4	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tfidf	False	True	dice	1,1,1,1	0.4497	0.3056	0.1754	0.0811	0.3103	0.3042	0.8856
0.6655											
tfidf	False	True	dice	3,3,4,1	0.4415	0.2997	0.1575	0.0628	0.2996	0.2884	0.8821
0.6523											
tfidf	False	True	dice	1,1,1,4	0.4770	0.2532	0.1457	0.0713	0.2920	0.2813	0.8818
0.6460											
tfidf	False	True	overlap	1,1,1,1	0.4541	0.3073	0.1750	0.0613	0.3121	0.3010	0.8891
0.6594											

tfidf	False	True	overlap	3,3,4,1	0.5138	0.3011	0.1868	0.0672	0.3339	0.3166	0.8867	0.6637
tfidf	False	True	overlap	1,1,1,4	0.4048	0.2679	0.1311	0.0263	0.2679	0.2599	0.8821	0.6244
tfidf	True	False	cosine	1,1,1,1	0.5719	0.3525	0.2074	0.0913	0.3772	0.3677	0.9331	0.7225
tfidf	True	False	cosine	3,3,4,1	0.5615	0.3601	0.2136	0.0776	0.3784	0.3622	0.9329	0.7237
tfidf	True	False	cosine	1,1,1,4	0.5127	0.2606	0.1567	0.0578	0.3100	0.3077	0.9245	0.6786
tfidf	True	False	jaccard	1,1,1,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796	0.1004
tfidf	True	False	jaccard	3,3,4,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796	0.1004
tfidf	True	False	jaccard	1,1,1,4	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796	0.1004
tfidf	True	False	dice	1,1,1,1	0.4785	0.2986	0.1917	0.0915	0.3229	0.3201	0.9305	0.7004
tfidf	True	False	dice	3,3,4,1	0.4930	0.3175	0.1826	0.0620	0.3310	0.3187	0.9292	0.6971
tfidf	True	False	dice	1,1,1,4	0.4372	0.2438	0.1552	0.0837	0.2787	0.2830	0.9208	0.6661
tfidf	True	False	overlap	1,1,1,1	0.4655	0.3187	0.1823	0.0589	0.3222	0.3118	0.9292	0.6896
tfidf	True	False	overlap	3,3,4,1	0.5121	0.3331	0.1909	0.0776	0.3454	0.3289	0.9329	0.7061
tfidf	True	False	overlap	1,1,1,4	0.3923	0.2449	0.1309	0.0299	0.2560	0.2517	0.9159	0.6370
tfidf	True	True	cosine	1,1,1,1	0.5918	0.3686	0.2232	0.1003	0.3945	0.3834	0.9348	0.7311
tfidf	True	True	cosine	3,3,4,1	0.5871	0.3863	0.2293	0.0913	0.4009	0.3815	0.9338	0.7307
tfidf	True	True	cosine	1,1,1,4	0.5170	0.2698	0.1624	0.0617	0.3164	0.3170	0.9270	0.6894
tfidf	True	True	jaccard	1,1,1,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796	0.1004

tfidf	True	True	jaccard	3,3,4,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tfidf	True	True	jaccard	1,1,1,4	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
tfidf	True	True	dice	1,1,1,1	0.5116	0.3300	0.2182	0.1126	0.3533	0.3473	0.9326
0.7146											
tfidf	True	True	dice	3,3,4,1	0.5001	0.3374	0.2013	0.0791	0.3462	0.3322	0.9304
0.7050											
tfidf	True	True	dice	1,1,1,4	0.4467	0.2509	0.1644	0.0868	0.2873	0.2887	0.9280
0.6806											
tfidf	True	True	overlap	1,1,1,1	0.4842	0.3376	0.2066	0.0787	0.3428	0.3313	0.9334
0.7064											
tfidf	True	True	overlap	3,3,4,1	0.5138	0.3439	0.2046	0.0872	0.3541	0.3376	0.9341
0.7130											
tfidf	True	True	overlap	1,1,1,4	0.4273	0.2641	0.1449	0.0380	0.2787	0.2711	0.9233
0.6597											
boolean	False	False	cosine	1,1,1,1	0.1627	0.0651	0.0291	0.0109	0.0856	0.0996	0.7816
0.4256											
boolean	False	False	cosine	3,3,4,1	0.2694	0.1066	0.0458	0.0073	0.1406	0.1474	0.8029
0.4708											
boolean	False	False	cosine	1,1,1,4	0.1401	0.0551	0.0288	0.0114	0.0747	0.0907	0.7702
0.4071											
boolean	False	False	jaccard	1,1,1,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
boolean	False	False	jaccard	3,3,4,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
boolean	False	False	jaccard	1,1,1,4	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
boolean	False	False	dice	1,1,1,1	0.1146	0.0422	0.0199	0.0064	0.0589	0.0757	0.7185
0.3396											
boolean	False	False	dice	3,3,4,1	0.1193	0.0387	0.0218	0.0062	0.0599	0.0764	0.7221
0.3393											
boolean	False	False	dice	1,1,1,4	0.1338	0.0472	0.0253	0.0093	0.0688	0.0922	0.7803
0.4086											
boolean	False	False	overlap	1,1,1,1	0.2560	0.1061	0.0546	0.0096	0.1389	0.1480	0.7982
0.4715											

boolean	False	False	overlap	3,3,4,1	0.3523	0.1276	0.0674	0.0082	0.1825	0.1828	0.8198
0.5158											
boolean	False	False	overlap	1,1,1,4	0.2340	0.1046	0.0551	0.0104	0.1313	0.1383	0.7897
0.4523											
boolean	False	True	cosine	1,1,1,1	0.1880	0.0841	0.0400	0.0093	0.1040	0.1203	0.7572
0.4307											
boolean	False	True	cosine	3,3,4,1	0.3080	0.1695	0.0645	0.0088	0.1807	0.1870	0.8089
0.5088											
boolean	False	True	cosine	1,1,1,4	0.1785	0.0789	0.0388	0.0099	0.0987	0.1141	0.7475
0.4147											
boolean	False	True	jaccard	1,1,1,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
boolean	False	True	jaccard	3,3,4,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
boolean	False	True	jaccard	1,1,1,4	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
boolean	False	True	dice	1,1,1,1	0.1401	0.0531	0.0281	0.0078	0.0738	0.0920	0.7162
0.3765											
boolean	False	True	dice	3,3,4,1	0.1528	0.0568	0.0304	0.0074	0.0800	0.0949	0.7279
0.3848											
boolean	False	True	dice	1,1,1,4	0.1843	0.0823	0.0455	0.0147	0.1040	0.1214	0.8280
0.4840											
boolean	False	True	overlap	1,1,1,1	0.2711	0.1260	0.0568	0.0118	0.1513	0.1602	0.8102
0.4857											
boolean	False	True	overlap	3,3,4,1	0.3894	0.2019	0.1163	0.0437	0.2358	0.2443	0.8474
0.5822											
boolean	False	True	overlap	1,1,1,4	0.2346	0.1024	0.0578	0.0137	0.1316	0.1413	0.8004
0.4608											
boolean	True	False	cosine	1,1,1,1	0.1809	0.0857	0.0354	0.0083	0.1007	0.1133	0.8197
0.4389											
boolean	True	False	cosine	3,3,4,1	0.2852	0.1235	0.0634	0.0093	0.1574	0.1671	0.8512
0.5077											
boolean	True	False	cosine	1,1,1,4	0.1758	0.0822	0.0355	0.0080	0.0978	0.1107	0.8096
0.4300											
boolean	True	False	jaccard	1,1,1,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											

boolean	True	False	jaccard	3,3,4,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
boolean	True	False	jaccard	1,1,1,4	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
boolean	True	False	dice	1,1,1,1	0.1486	0.0633	0.0279	0.0069	0.0800	0.0928	0.7475
0.3524											
boolean	True	False	dice	3,3,4,1	0.1463	0.0600	0.0223	0.0068	0.0762	0.0891	0.7644
0.3627											
boolean	True	False	dice	1,1,1,4	0.1469	0.0512	0.0274	0.0114	0.0751	0.0967	0.8098
0.4186											
boolean	True	False	overlap	1,1,1,1	0.2219	0.0861	0.0528	0.0094	0.1203	0.1381	0.8313
0.4828											
boolean	True	False	overlap	3,3,4,1	0.3707	0.1270	0.0517	0.0122	0.1832	0.1921	0.8618
0.5462											
boolean	True	False	overlap	1,1,1,4	0.2328	0.0819	0.0497	0.0093	0.1215	0.1330	0.8216
0.4650											
boolean	True	True	cosine	1,1,1,1	0.2182	0.0982	0.0437	0.0076	0.1200	0.1318	0.8047
0.4491											
boolean	True	True	cosine	3,3,4,1	0.3154	0.1550	0.0713	0.0112	0.1806	0.1909	0.8646
0.5441											
boolean	True	True	cosine	1,1,1,4	0.2233	0.1035	0.0435	0.0077	0.1234	0.1362	0.7958
0.4426											
boolean	True	True	jaccard	1,1,1,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
boolean	True	True	jaccard	3,3,4,1	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
boolean	True	True	jaccard	1,1,1,4	0.0780	0.0182	0.0115	0.0044	0.0359	0.0526	0.2796
0.1004											
boolean	True	True	dice	1,1,1,1	0.1722	0.0700	0.0362	0.0073	0.0928	0.1085	0.7520
0.3855											
boolean	True	True	dice	3,3,4,1	0.1676	0.0638	0.0346	0.0081	0.0887	0.1034	0.7767
0.4010											
boolean	True	True	dice	1,1,1,4	0.2106	0.0849	0.0378	0.0106	0.1111	0.1270	0.8633
0.4865											
boolean	True	True	overlap	1,1,1,1	0.2715	0.1120	0.0628	0.0103	0.1488	0.1643	0.8533
0.5119											

boolean	True	True	overlap	3,3,4,1	0.4545	0.2153	0.1134	0.0469	0.2611	0.2601	0.8919 0.6110
boolean	True	True	overlap	1,1,1,4	0.2661	0.1013	0.0563	0.0102	0.1412	0.1540	0.8444 0.4897

2. **List the top 20 retrieved documents for Queries 6, 9 and 22 by their number, title and similarity measure, with the "relevant" documents starred.**

Query 6: Stem = True Similarity = cosine

query_id doc_id similarity title relevant

*6 1543 0.21152517770032234 computer formulation of the equations of motion using tensor notation True
*6 2078 0.18771975208100086 representations for space planning True
6 136 0.16151115965448135 a note on the calculation of interest False
*6 2828 0.139906178345471 hierarchical geometric models for visible surface algorithms True
6 356 0.13789226102062566 interest (algorithm 45) False
6 740 0.1372398389378782 interest (algorithm 45) False
6 242 0.12423915393758846 notes on geometric weighted check digit verification False
6 2389 0.11159810681461133 preliminary report on a system for general space planning False
6 1186 0.10958847087681071 recursive solution of a class of combinatorial problems : an example False
6 1398 0.10706445623639943 robot data screening : a solution to multivariate type problems in the biological and social sciences False
6 3035 0.10483224784002099 a strategic planning methodology for the computing effort in higher education : an empirical evaluation False
6 2671 0.10387799431824886 a note on a combinatorial problem of burnett and coffman False
6 1009 0.10254395395301849 solution of combinatorial problems using generating functions on a variable-field computer False
6 2230 0.09955439496420586 a language for treating geometric patterns in a two-dimensional space False
6 1755 0.09947094788662406 proceedings of the acm symposium on operating system principles False
6 2826 0.09763016759302308 interactive skeleton techniques for enhancing motion dynamics in key frame animation False
6 2187 0.09479111940719465 computer science : a conceptual framework for curriculum planning False
6 2753 0.09253702363080796 a heuristic problem solving design system for equipment or furniture layouts False
6 705 0.08148842063409324 combinatorial of m things taken n at a time (algorithm 160) False
6 1517 0.07932051561864047 methods for analyzing data from computer simulation experiments False

Query 9: Stem = True Similarity = cosine

9 1685 0.5285390575627126 gan , a system for generating and analyzing activity networks False
9 2949 0.4278639240339402 a correctness proof of a topology information main tenance protocol for a distributed computer network False
*9 3158 0.38525375967677317 secure personal computing in an insecure network True
9 2197 0.38294334978022976 the merit of regional computing networks False
9 2776 0.3708980061254136 computer networks in higher education : socio-economic-political factors False
9 2614 0.33779929220271177 arrow to precedence network transformation [h] (algorithm a481) False
9 2951 0.335059232177022 dynamic response time prediction for computer networks False
*9 3111 0.33208038709881027 secure communications over insecure channels True
*9 3068 0.32124061706687107 a model for verification of data security in operating systems True
9 2969 0.31345202437837083 optimal program and data locations in computer networks False
9 1695 0.30236090741215593 plexus-an on-line system for modeling neural networks False
9 2864 0.29758236776338287 characteristics of program localities False
9 1261 0.29582303511169505 modeling and simulation of digital networks False
*9 2372 0.29074902694311494 on the implementation of security measures in information systems True
9 1723 0.2901449112923111 computer construction of project networks False
9 2515 0.28797063721151966 minimal event-node network of project precedence relations False
9 2454 0.28459010918494476 computational algorithms for closed queueing networks with exponential servers False
9 2371 0.28325379114135196 a system for interprocess communication in a resource sharing computer network False
*9 2870 0.2784856872671307 a lattice model of secure information flow True
9 1611 0.2647311125845895 scheduling project networks False

Query 22: Stem = True Similarity = cosine

22 266 0.17991968117850357 fitting spheres by the method of least squares False
*22 2473 0.17688143334899786 hidden-line plotting program (algorithm r420) True
*22 2384 0.1761342839872564 hidden-line plotting program [j6] (algorithm a420) True
*22 2441 0.16228583502259314 hidden-line plotting program (algorithm r420) True
*22 2638 0.15719761209897773 hidden-line plotting program (algorithm r420) True

*22 2564 0.15698767733416613 hidden-line plotting program (algorithm r420) True
*22 2637 0.15698767733416613 hidden-line plotting program (algorithm r420) True
22 2913 0.15597673867434256 the aliasing problem in computer-generated shaded images False
22 3049 0.14994360253282044 a simply extended and modified batch environment graphical system (sembegs) False
22 87 0.14874181120398794 a note on a method for generating points uniformly on n-dimensional spheres False
22 122 0.13688243830638266 least squares fitting of a great circle through points on a sphere False
22 52 0.12779984576811218 an efficient method for generating uniformly distributed points on the surface on an n-dimensional sphere (corrigendum) False
22 1829 0.12615680967971246 an interactive graphical display monitor in a batch-processing environment with remote entry False
22 2809 0.12372552370490981 positivity and norms False
22 88 0.11941264554931032 an efficient method for generating uniformly distributed points on the surface of an n-dimensional sphere False
22 1978 0.11820776904507375 the use of interactive graphics to solve numerical problems False
22 2105 0.10534010076095673 an interactive computer system using graphical flowchart input False
22 1467 0.1019692770663841 a graphical servicesystem with variable syntax False
22 1466 0.09942074340675659 discussion summary on graphical languages False
22 1767 0.09915500412025706 a general purpose graphic language False

Query 6: Stem = False Similarity = cosine

*6 1543 0.1919382921470839 computer formulation of the equations of motion using tensor notation True
*6 2828 0.1336044186271193 hierarchical geometric models for visible surface algorithms True
6 2826 0.12455746758138564 interactive skeleton techniques for enhancing motion dynamics in key frame animation False
6 2389 0.11571600970848182 preliminary report on a system for general space planning False
6 242 0.11421398229664403 notes on geometric weighted check digit verification False
6 3035 0.10803165362214541 a strategic planning methodology for the computing effort in higher education : an empirical evaluation False
6 2187 0.10003308158358365 computer science : a conceptual framework for curriculum planning False
6 2230 0.09633169144121556 a language for treating geometric patterns in a two-dimensional space False

6 2753 0.09608068527256108 a heuristic problem solving design system for equipment or furniture layouts False
6 2721 0.09398474219446869 the digital simulation of river plankton population dynamics False
6 1186 0.09169093358461779 recursive solution of a class of combinatorial problems : an example False
6 2671 0.08786023503704705 a note on a combinatorial problem of burnett and coffman False
6 1009 0.0838124738102087 solution of combinatorial problems using generating functions on a variable-field computer False
6 530 0.07254427320802072 a computer method for radiation treatment planning False
6 705 0.07242646375247062 combinatorial of m things taken n at a time (algorithm 160) False
6 704 0.06963343596402154 combinatorial of m things taken one at a time two at a time , up to n at a time (algorithm 161) False
6 888 0.06748140317278385 algorithm 160 combinatorial of m things taken n at a time False
6 2087 0.06490222770580117 a number system for the permutations False
6 695 0.06267398781483877 use of the disk file on stretch False
6 1014 0.06265347740803998 machine controls for analysis of variance False

Query 9: Stem = False Similarity = cosine

9 2949 0.3524003658656976 a correctness proof of a topology information main tenance protocol for a distributed computer network False
*9 3068 0.33489081407219945 a model for verification of data security in operating systems True
*9 2372 0.3116546012971935 on the implementation of security measures in information systems True
9 3082 0.30002150954647894 time , clocks , and the ordering of events in a distributed system False
9 3174 0.29960805795800105 password security : a case history False
9 1685 0.2945588711002893 gan , a system for generating and analyzing activity networks False
9 2621 0.28759317523660893 a high security log-in procedure False
9 2849 0.2848478088050912 ethernet : distributed packet switching for local computer networks False
9 3137 0.23863028730474345 a methodology for the design of distributed information systems False
9 2969 0.2313403360271588 optimal program and data locations in computer networks False
9 2776 0.2265353441947695 computer networks in higher education : socio-economic-political factors False
9 3148 0.21933987108785133 high level programming for distributed computing False
*9 3158 0.2147718434739509 secure personal computing in an insecure network True

9 1750 0.21474894446573106 considerations in the design of a multiple computer system with extended core storage False
*9 3111 0.21038930249726673 secure communications over insecure channels True
9 2614 0.20479105551584578 arrow to precedence network transformation [h] (algorithm a481) False
9 2900 0.2015109597512051 some theorems to aid in solving the file allocation problem False
9 1461 0.19947946654863585 discussion summary on operating systems False
9 2578 0.19747919076027634 self-stabilizing systems in spite of distributed control False
*9 2870 0.19345619217259508 a lattice model of secure information flow True

Query 22: Stem = False Similarity = cosine

22 266 0.22124292793944447 fitting spheres by the method of least squares False
22 87 0.17865108474437166 a note on a method for generating points uniformly on n-dimensional spheres False
*22 2473 0.16846638471863648 hidden-line plotting program (algorithm r420) True
*22 2384 0.16773377567396117 hidden-line plotting program [j6] (algorithm a420) True
22 2913 0.150377715129546 the aliasing problem in computer-generated shaded images False
*22 2441 0.14805812756427905 hidden-line plotting program (algorithm r420) True
*22 2638 0.14353839149515285 hidden-line plotting program (algorithm r420) True
*22 2564 0.14335166467173693 hidden-line plotting program (algorithm r420) True
*22 2637 0.14335166467173693 hidden-line plotting program (algorithm r420) True
22 1978 0.11680615294496183 the use of interactive graphics to solve numerical problems False
*22 2692 0.11104536824091366 reentrant polygon clipping True
22 2809 0.09018716484174505 positivity and norms False
22 2924 0.08215103608579259 an interactive computer graphics approach to surface representation False
*22 2369 0.07722881828526763 hidden lines elimination for a rotating object True
22 2211 0.0722763747602631 scanned-display computer graphics False
22 2188 0.06900139254471395 an approach to the optimum design of computer graphics systems False
22 3049 0.06527434027379941 a simply extended and modified batch environment graphical system (sembegs) False
22 2152 0.06116564849579829 display procedures False
22 2674 0.05998106631543988 scan conversion algorithms for a cell organized raster display False
22 2004 0.058898018846943866 a procedure for generation of three-dimensional half-toned computer graphics presentations False

2. For the top 10 retrieved documents, show the terms on which the retrieval was based (those with non-zero weights for both query and retrieved document) along with these weights.

Note: In order to shorten the output, I only included output for query 6,9,22. Complete output can be produced by including the whole query.

Query 6: Stem = True Similarity = cosine

query_id	doc_id	terms	doc_weight	query_weight
6	1543	,	0.18348183074072794	0.016680166430975266
6	1543	.	0.22728251077556091	0.04545650215511218
6	1543	motion	28.371300176949394	11.348520070779758
6	2078	,	0.45036449363633220	.016680166430975266
6	2078	.	0.13636950646533655	0.04545650215511218
6	2078	robot	29.5160325105132177	.379008127628304
6	2078	plan	22.7289739178604444	.545794783572089
6	136	,	0.05004049929292580	.016680166430975266
6	136	.	0.06818475323266827	0.04545650215511218
6	136	interest	11.8350627694294757	.890041846286317
6	2828	,	0.41700416077438160	.016680166430975266
6	2828	.	0.22728251077556091	0.04545650215511218
6	2828	geometr	55.8724865840024955	.58724865840025
6	2828	motion	5.674260035389879	11.348520070779758
6	356	,	0.05004049929292580	.016680166430975266
6	356	.	0.06818475323266827	0.04545650215511218
6	356	interest	11.8350627694294757	.890041846286317
6	740	,	0.05004049929292580	.016680166430975266
6	740	.	0.06818475323266827	0.04545650215511218
6	740	interest	11.8350627694294757	.890041846286317
6	242	,	0.05004049929292580	.016680166430975266
6	242	.	0.11364125538778046	0.04545650215511218
6	242	geometr	22.348994633601	5.58724865840025
6	2389	,	0.28356282932657950	.016680166430975266
6	2389	.	0.15909775754289263	0.04545650215511218
6	2389	plan	36.36635826857671	4.545794783572089
6	1186	,	0.08340083215487633	0.016680166430975266
6	1186	.	0.18182600862044873	0.04545650215511218
6	1186	combinatori	18.9598031920521854	.739950798013046
6	1398	,	0.20016199717170320	.016680166430975266
6	1398	.	0.36365201724089746	0.04545650215511218
6	1398	robot	22.1370243828849137	.379008127628304

Query 9: Stem = True Similarity = cosine

9	1685	,	0.6171661579460849	0.03336033286195053
9	1685	.	0.13636950646533655	0.02272825107755609
9	1685	network	137.037393202279337	.212494379067333
9	1685	system	6.016309587105097	3.0081547935525483
9	2949	,	0.6838682366998590	.03336033286195053
9	2949	.	0.20455425969800484	0.02272825107755609
9	2949	distribut	53.51627983161455	3.344767489475909
9	2949	network	82.94368535927433	7.212494379067333
9	2949	oper	8.613045815660413	2.1532614539151034
9	2949	system	6.016309587105097	3.0081547935525483
9	3158	,	0.31692316218853006	0.03336033286195053
9	3158	.	0.15909775754289263	0.02272825107755609
9	3158	secur	46.70957926526094	4.6709579265260945
9	3158	network	32.4562247058029957	.212494379067333
9	2197	,	0.5838058250841343	0.03336033286195053
9	2197	.	0.22728251077556091	0.02272825107755609
9	2197	network	75.73119098020699	7.212494379067333
9	2197	oper	2.1532614539151034	2.1532614539151034
9	2776	,	0.2835628293265795	0.03336033286195053
9	2776	.	0.34092376616334136	0.02272825107755609
9	2776	network	57.69995503253866	7.212494379067333
9	2776	oper	2.1532614539151034	2.1532614539151034
9	2776	consider	3.8977680382926128	3.8977680382926128
9	2614	,	0.18348183074072794	0.03336033286195053
9	2614	.	0.06818475323266827	0.02272825107755609
9	2614	network	39.66871908487033	7.212494379067333
9	2951	,	0.6672066572390106	0.03336033286195053
9	2951	.	0.15909775754289263	0.02272825107755609
9	2951	network	61.30620222207233	7.212494379067333
9	2951	system	18.04892876131529	3.0081547935525483
9	3111	,	0.6505264908080354	0.03336033286195053
9	3111	.	0.18182600862044873	0.02272825107755609
9	3111	secur	84.0772426774697	4.6709579265260945
9	3111	network	14.4249887581346657	.212494379067333
9	3111	distribut	13.3790699579036373	.344767489475909
9	3068	,	0.31692316218853006	0.03336033286195053
9	3068	.	0.15909775754289263	0.02272825107755609
9	3068	oper	21.532614539151034	2.1532614539151034
9	3068	system	16.5448513645390173	.0081547935525483
9	3068	secur	51.38053719178704	4.6709579265260945

9	2969	,	0.31692316218853006	0.03336033286195053
9	2969	.	0.22728251077556091	0.02272825107755609
9	2969	network	32.4562247058029957	.212494379067333
9	2969	distribut	13.3790699579036373	.344767489475909

Query 22: Stem = True Similarity = cosine

22	266	,	0.05004049929292580	.0500404992929258
22	266	.	0.06818475323266827	0.04545650215511218
22	266	sphere	18.3787354773988136	.126245159132937
22	2473	,	0.16680166430975266	0.0500404992929258
22	2473	.	0.13636950646533655	0.04545650215511218
22	2473	hidden-lin	43.9627708727213646	.280395838960195
22	2473	algorithm	2.630903964028623	0.8769679880095411
22	2384	,	0.11676116501682686	0.0500404992929258
22	2384	.	0.06818475323266827	0.04545650215511218
22	2384	hidden-lin	43.9627708727213646	.280395838960195
22	2384	algorithm	2.630903964028623	0.8769679880095411
22	2441	,	0.05004049929292580	.0500404992929258
22	2441	.	0.06818475323266827	0.04545650215511218
22	2441	hidden-lin	18.8411875168805846	.280395838960195
22	2441	algorithm	2.630903964028623	0.8769679880095411
22	2638	,	0.05004049929292580	.0500404992929258
22	2638	.	0.06818475323266827	0.04545650215511218
22	2638	hidden-lin	18.8411875168805846	.280395838960195
22	2638	algorithm	2.630903964028623	0.8769679880095411
22	2564	,	0.05004049929292580	.0500404992929258
22	2564	.	0.06818475323266827	0.04545650215511218
22	2564	hidden-lin	18.8411875168805846	.280395838960195
22	2564	algorithm	2.630903964028623	0.8769679880095411
22	2637	,	0.05004049929292580	.0500404992929258
22	2637	.	0.06818475323266827	0.04545650215511218
22	2637	hidden-lin	18.8411875168805846	.280395838960195
22	2637	algorithm	2.630903964028623	0.8769679880095411
22	2913	,	0.40032399434340640	.0500404992929258
22	2913	.	0.15909775754289263	0.04545650215511218
22	2913	comput	5.030449643713166	1.2576124109282916
22	2913	graphic	14.861785925994633	.7154464814986583
22	2913	hidden-surfac	36.89504063814152	7.379008127628304
22	2913	algorithm	0.8769679880095411	0.8769679880095411
22	3049	,	0.567125658653159	0.0500404992929258
22	3049	.	0.20455425969800484	0.04545650215511218

22	3049	comput	6.288062054641458	1.2576124109282916
22	3049	graphic	133.7560733339517	3.7154464814986583
22	87	,	0.05004049929292580	.0500404992929258
22	87	.	0.06818475323266827	0.04545650215511218
22	87	sphere	18.3787354773988136	.126245159132937

Query 6: Stem = False Similarity = cosine

query_id	doc_id	terms	doc_weight	query_weight
6	1543	,	0.18348183074072794	0.016680166430975266
6	1543	.	0.22728251077556091	0.04545650215511218
6	1543	motion	29.963568832542073	11.985427533016829
6	2828	,	0.41700416077438160	.016680166430975266
6	2828	.	0.22728251077556091	0.04545650215511218
6	2828	geometric	59.927137665084146	5.992713766508414
6	2828	motion	5.992713766508414	11.985427533016829
6	2826	,	0.36696366148145587	0.016680166430975266
6	2826	.	0.250010761853117	0.04545650215511218
6	2826	motion	29.963568832542073	11.985427533016829
6	2826	dynamics	26.74344378827344	6.68586094706836
6	2389	,	0.28356282932657950	.016680166430975266
6	2389	.	0.15909775754289263	0.04545650215511218
6	2389	planning	40.61138427707407	5.076423034634259
6	242	,	0.05004049929292580	.016680166430975266
6	242	.	0.11364125538778046	0.04545650215511218
6	242	geometric	23.970855066033657	5.992713766508414
6	3035	,	0.35028349505048056	0.016680166430975266
6	3035	.	0.250010761853117	0.04545650215511218
6	3035	planning	55.84065338097685	5.076423034634259
6	2187	,	0.28356282932657950	.016680166430975266
6	2187	.	0.13636950646533655	0.04545650215511218
6	2187	planning	40.61138427707407	5.076423034634259
6	2230	,	0.750607489393887	0.016680166430975266
6	2230	.	0.20455425969800484	0.04545650215511218
6	2230	planning	25.382115173171297	5.076423034634259
6	2230	geometric	23.970855066033657	5.992713766508414
6	2753	,	0.93408932013461490	.016680166430975266
6	2753	.	0.34092376616334136	0.04545650215511218
6	2753	planning	65.99349945024537	5.076423034634259
6	2721	,	0.91740915370363960	.016680166430975266
6	2721	.	0.18182600862044873	0.04545650215511218
6	2721	dynamics	53.48688757654688	6.68586094706836

6 2721 planning 25.3821151731712975.076423034634259

Query 9: Stem = False Similarity = cosine

9 2949 , 0.6838868236699859 0.03336033286195053
9 2949 . 0.20455425969800484 0.02272825107755609
9 2949 distributed 76.42109507494273 4.77631844218392
9 2949 network 58.01194033195926 3.867462688797284
9 2949 networks 33.77606165182553 4.222007706478191
9 2949 operating 13.03988381126333 3.2599709528158325
9 3068 , 0.31692316218853006 0.03336033286195053
9 3068 . 0.15909775754289263 0.02272825107755609
9 3068 operating 32.59970952815833 3.2599709528158325
9 3068 systems 19.04686144851077 4.232635877446838
9 3068 security 42.03862133873485 4.6709579265260945
9 2372 , 0.567125658653159 0.03336033286195053
9 2372 . 0.29546726400822920.02272825107755609
9 2372 security 51.38053719178704 4.6709579265260945
9 2372 operating 16.299854764079164 3.2599709528158325
9 2372 systems 27.5121332034044464.232635877446838
9 3082 , 0.38364382791243110.03336033286195053
9 3082 . 0.15909775754289263 0.02272825107755609
9 3082 distributed 42.986865979655285 4.77631844218392
9 3082 systems 16.9305435097873524.232635877446838
9 3082 networks 16.888030825912764 4.222007706478191
9 3174 , 0.23352233003365372 0.03336033286195053
9 3174 . 0.20455425969800484 0.02272825107755609
9 3174 operating 13.03988381126333 3.2599709528158325
9 3174 systems 8.465271754893676 4.232635877446838
9 3174 security 42.03862133873485 4.6709579265260945
9 1685 , 0.61716615794608490.03336033286195053
9 1685 . 0.13636950646533655 0.02272825107755609
9 1685 network 116.023880663918523.867462688797284
9 1685 networks 33.77606165182553 4.222007706478191
9 2621 , 0.26688266289560425 0.03336033286195053
9 2621 . 0.13636950646533655 0.02272825107755609
9 2621 operating 13.03988381126333 3.2599709528158325
9 2621 systems 19.04686144851077 4.232635877446838
9 2621 security 32.69670548568266 4.6709579265260945
9 2849 , 0.55044549222218380.03336033286195053
9 2849 . 0.29546726400822920.02272825107755609

9	2849	networks	33.77606165182553	4.222007706478191
9	2849	distributed	66.86845819057488	4.77631844218392
9	2849	local	19.9244514193197344	4.9811128548299335
9	2849	systems	2.116317938723419	4.232635877446838
9	2849	operating	3.2599709528158325	3.2599709528158325
9	3137	,	0.500404992929258	0.03336033286195053
9	3137	.	0.22728251077556091	0.02272825107755609
9	3137	distributed	57.31582130620704	4.77631844218392
9	3137	systems	6.348953816170257	4.232635877446838
9	3137	operating	3.2599709528158325	3.2599709528158325
9	2969	,	0.31692316218853006	0.03336033286195053
9	2969	.	0.22728251077556091	0.02272825107755609
9	2969	networks	33.77606165182553	4.222007706478191
9	2969	distributed	19.10527376873568	4.77631844218392
9	2969	network	3.867462688797284	3.867462688797284

Query 22: Stem = False Similarity = cosine

22	266	,	0.05004049929292580	.0500404992929258
22	266	.	0.06818475323266827	0.04545650215511218
22	266	spheres	22.1370243828849137	.379008127628304
22	87	,	0.05004049929292580	.0500404992929258
22	87	.	0.06818475323266827	0.04545650215511218
22	87	spheres	22.1370243828849137	.379008127628304
22	2473	,	0.16680166430975266	0.0500404992929258
22	2473	.	0.13636950646533655	0.04545650215511218
22	2473	hidden-line	43.962770872721364	6.280395838960195
22	2384	,	0.11676116501682686	0.0500404992929258
22	2384	.	0.06818475323266827	0.04545650215511218
22	2384	hidden-line	43.962770872721364	6.280395838960195
22	2913	,	0.40032399434340640	.0500404992929258
22	2913	.	0.15909775754289263	0.04545650215511218
22	2913	computer	6.556860861796281	1.6392152154490702
22	2913	graphics	16.721340040310494	4.1803350100776235
22	2913	hidden-surface	36.89504063814152	7.379008127628304
22	2913	algorithms	2.6968769005040847	2.6968769005040847
22	2441	,	0.05004049929292580	.0500404992929258
22	2441	.	0.06818475323266827	0.04545650215511218
22	2441	hidden-line	18.841187516880584	6.280395838960195
22	2638	,	0.05004049929292580	.0500404992929258
22	2638	.	0.06818475323266827	0.04545650215511218
22	2638	hidden-line	18.841187516880584	6.280395838960195
22	2564	,	0.05004049929292580	.0500404992929258

22	2564	.	0.06818475323266827	0.04545650215511218
22	2564	hidden-line	18.8411875168805846	.280395838960195
22	2637	,	0.0500404992929258	0.0500404992929258
22	2637	.	0.06818475323266827	0.04545650215511218
22	2637	hidden-line	18.8411875168805846	.280395838960195
22	1978	,	0.5504454922221838	0.0500404992929258
22	1978	.	0.22728251077556091	0.04545650215511218
22	1978	graphics	62.70502515116435	4.1803350100776235
22	1978	computer	8.196076077245351	1.6392152154490702

3. **List the top 20 documents that are most similar to Documents 239, 1236 and 2740, giving number, title and similarity measure.**

Query 239 : Stem = True Similarity = cosine

query_id doc_id similarity title
239 1032 0.5139802019059281 theoretical considerations in information retrieval systems
239 2965 0.19325371214505924 an optimal evaluation of boolean expressions in an online query system
239 2160 0.18734701947846777 canonical structure in attribute based file organization
239 3168 0.17619198586214607 comment on `` an optimal evaluation of boolean expressions in an online query system . ''
239 3169 0.1671346590770386 note on `` an optimal evaluation of boolean expressions in an online query system . ''
239 1207 0.1538093423604967 remarks on simulation of boolean functions
239 1329 0.15267103432714313 simulation of boolean functions in a decimal computer
239 3134 0.1378387479970241 the use of normal multiplication tables for information storage and retrieval
239 651 0.13658143368700623 a survey of languages and systems for information retrieval
239 1457 0.1309287829664454 data manipulation and programming problems in automatic information retrieval
239 2345 0.12844643437245498 curriculum recommendations for graduate professional programs in information systems
239 891 0.12118502082473853 everyman 's information retrieval system
239 2824 0.11994074072822801 an improvement to martin 's algorithm for computation of linear precedence functions
239 3012 0.11460679268222612 the use of an interactive information storage and retrieval system in medical research
239 1699 0.11433081829452611 experimental evaluation of information retrieval through a teletypewriter
239 2278 0.11316487895455132 on foster 's information storage and retrieval using avl trees
239 2340 0.11084435750566538 a boolean matrix method for the computation of linear precedence functions

239 1927 0.10693206636615744 information science in a ph.d. computer science program
239 2516 0.10525881681952966 hierarchical storage in information retrieval
239 2479 0.10358014648209501 curriculum recommendations for graduate professional programs in information systems : recommended addendum on information systems administration

Query 1236 : Stem = True Similarity = cosine

1236 2711 0.2829728252638341 a vector space model for automatic indexing
1236 2307 0.2829566119255625 dynamic document processing
1236 1457 0.2776816846614699 data manipulation and programming problems in automatic information retrieval
1236 1699 0.23237128998756035 experimental evaluation of information retrieval through a teletypewriter
1236 2575 0.20859606143234322 the best-match problem in document retrieval
1236 2990 0.16518456583203395 effective information retrieval using term accuracy
1236 1681 0.1565964016898641 easy english , a language for information retrieval through a remote typewriter console
1236 634 0.1497870654237051 manipulation of trees in information retrieval*
1236 3012 0.13183286419903878 the use of an interactive information storage and retrieval system in medical research
1236 3134 0.1264421885857958 the use of normal multiplication tables for information storage and retrieval
1236 891 0.12193253623983608 everyman 's information retrieval system
1236 1927 0.12005457114404916 information science in a ph.d. computer science program
1236 2278 0.11442744177797462 on foster 's information storage and retrieval using avl trees
1236 1935 0.11302386789880108 randomized binary search technique
1236 3135 0.10964440710572047 detection of three-dimensional patterns of atoms in chemical structures
1236 1536 0.1095920783603435 dynamic computation of derivatives
1236 3084 0.1092086043384976 interpolation search -a log logn search
1236 2947 0.10777561574169961 sitar : an interactive text processing system for small computers
1236 1937 0.10487783162177125 codas : a data display system
1236 329 0.10310860129060806 automatic abstracting and indexing survey and recommendations

Query 2740 : Stem = True Similarity = cosine

2740 1749 0.33819452173415726 the structure of the `` the '' -multiprogramming system
2740 2379 0.27285597922755783 the design of the venus operating system

2740 2378 0.23359140909586207 an operating system based on the concept of a supervisory computer
2740 2920 0.22695683444157147 game interpretation of the deadlock avoidance problem
2740 2597 0.21600358716785165 monitors : an operating system structuring concept
2740 2342 0.20573972876952082 interference between communicating parallel processes
2740 2228 0.19616934491520183 comments on prevention of system deadlocks
2740 3043 0.19596149508936875 distributed processes : a concurrent programming concept
2740 2500 0.19319794930183704 a practical approach to managing resources and avoiding deadlocks
2740 2280 0.18788732726345753 comment on deadlock prevention method
2740 2700 0.1867181516294606 reduction : a method of proving properties of parallel programs
2740 2796 0.18304507811063134 monitors : an operating system structuring concept (corrigendum)
2740 1611 0.1799591993532193 scheduling project networks
2740 2080 0.1792192905843657 the nucleus of a multiprogramming system
2740 2376 0.16744751228728735 synchronization of communicating processes
2740 2320 0.16233646859289647 structured multiprogramming
2740 2482 0.1618551727980061 mixed solutions for the deadlock problem
2740 2542 0.15845595893832062 a software design and evaluation system
2740 2865 0.1550811314202658 verifying properties of parallel programs : an axiomatic approach
2740 1723 0.1526460416242082 computer construction of project networks

Query 239 : Stem = False Similarity = cosine

query_id doc_id similarity title
239 1032 0.5139802019059281 theoretical considerations in information retrieval systems
239 2965 0.19325371214505924 an optimal evaluation of boolean expressions in an online query system
239 2160 0.18734701947846777 canonical structure in attribute based file organization
239 3168 0.17619198586214607 comment on `` an optimal evaluation of boolean expressions in an online query system . "
239 3169 0.1671346590770386 note on `` an optimal evaluation of boolean expressions in an online query system . "
239 1207 0.1538093423604967 remarks on simulation of boolean functions
239 1329 0.15267103432714313 simulation of boolean functions in a decimal computer
239 3134 0.1378387479970241 the use of normal multiplication tables for information storage and retrieval
239 651 0.13658143368700623 a survey of languages and systems for information retrieval
239 1457 0.1309287829664454 data manipulation and programming problems in automatic information retrieval

239 2345 0.12844643437245498 curriculum recommendations for graduate professional programs in information systems
239 891 0.12118502082473853 everyman 's information retrieval system
239 2824 0.11994074072822801 an improvement to martin 's algorithm for computation of linear precedence functions
239 3012 0.11460679268222612 the use of an interactive information storage and retrieval system in medical research
239 1699 0.11433081829452611 experimental evaluation of information retrieval through a teletypewriter
239 2278 0.11316487895455132 on foster 's information storage and retrieval using avl trees
239 2340 0.11084435750566538 a boolean matrix method for the computation of linear precedence functions
239 1927 0.10693206636615744 information science in a ph.d. computer science program
239 2516 0.10525881681952966 hierarchical storage in information retrieval
239 2479 0.10358014648209501 curriculum recommendations for graduate professional programs in information systems : recommended addendum on information systems administration

Query 1236 : Stem = False Similarity = cosine

1236 2711 0.2829728252638341 a vector space model for automatic indexing
1236 2307 0.2829566119255625 dynamic document processing
1236 1457 0.2776816846614699 data manipulation and programming problems in automatic information retrieval
1236 1699 0.23237128998756035 experimental evaluation of information retrieval through a teletypewriter
1236 2575 0.20859606143234322 the best-match problem in document retrieval
1236 2990 0.16518456583203395 effective information retrieval using term accuracy
1236 1681 0.1565964016898641 easy english , a language for information retrieval through a remote typewriter console
1236 634 0.1497870654237051 manipulation of trees in information retrieval*
1236 3012 0.13183286419903878 the use of an interactive information storage and retrieval system in medical research
1236 3134 0.1264421885857958 the use of normal multiplication tables for information storage and retrieval
1236 891 0.12193253623983608 everyman 's information retrieval system
1236 1927 0.12005457114404916 information science in a ph.d. computer science program
1236 2278 0.11442744177797462 on foster 's information storage and retrieval using avl trees
1236 1935 0.11302386789880108 randomized binary search technique
1236 3135 0.10964440710572047 detection of three-dimensional patterns of atoms in chemical structures

1236 1536 0.1095920783603435 dynamic computation of derivatives
1236 3084 0.1092086043384976 interpolation search -a log logn search
1236 2947 0.10777561574169961 sitar : an interactive text processing system for small computers
1236 1937 0.10487783162177125 codas : a data display system
1236 329 0.10310860129060806 automatic abstracting and indexing survey and recommendations

Query 2740 : Stem = True Similarity = cosine

2740 1749 0.33819452173415726 the structure of the `` the '' -multiprogramming system
2740 2379 0.27285597922755783 the design of the venus operating system
2740 2378 0.23359140909586207 an operating system based on the concept of a supervisory computer
2740 2920 0.22695683444157147 game interpretation of the deadlock avoidance problem
2740 2597 0.21600358716785165 monitors : an operating system structuring concept
2740 2342 0.20573972876952082 interference between communicating parallel processes
2740 2228 0.19616934491520183 comments on prevention of system deadlocks
2740 3043 0.19596149508936875 distributed processes : a concurrent programming concept
2740 2500 0.19319794930183704 a practical approach to managing resources and avoiding deadlocks
2740 2280 0.18788732726345753 comment on deadlock prevention method
2740 2700 0.1867181516294606 reduction : a method of proving properties of parallel programs
2740 2796 0.18304507811063134 monitors : an operating system structuring concept (corrigendum)
2740 1611 0.1799591993532193 scheduling project networks
2740 2080 0.1792192905843657 the nucleus of a multiprogramming system
2740 2376 0.16744751228728735 synchronization of communicating processes
2740 2320 0.16233646859289647 structured multiprogramming
2740 2482 0.1618551727980061 mixed solutions for the deadlock problem
2740 2542 0.15845595893832062 a software design and evaluation system
2740 2865 0.1550811314202658 verifying properties of parallel programs : an axiomatic approach
2740 1723 0.1526460416242082 computer construction of project networks

Writeup
By conducting experiments with different permutations, I have noticed the following observations and ideas with different conditions:
Tfidf is the best weighting system. The tf weighting and boolean weighting are very similar, so term frequency is not the most crucial factor in building a retrieval system. More needs to be considered other than term frequency to build a good weighting system. Stemming the text

improves the results by a little. Stopwords can sometimes reduce the precision of the results. The region weightings performed better than default weighting, but since a lot of the queries/doc are missing with some parts, it does not have as strong impact for this particular set of data. Thus, we need to examine our queries and datasets clearly before building a good retrieval system in the future.

Part 3- Extensions to the Retrieval Model

The first extension is extension number 2. I created a dataset of size 100 documents called news1.raw for different news and articles, and 10 queries in news_query.raw regarding information about news. I have also listed all the relevant documents in news_query.rels.
The second extension I implemented is extension number 10, where I can accept queries directly from keyboard, and retrieve the top 20 most relevant documents. It will first prompt the user to choose to quit or continue with query, and then the user could choose to enter queries into the categories of author, title, keyword and abstract, and the user can continue to enter queries until he/she chooses to stop by pressing ENTER.