

Vamos a comentar una cosa antes de nada, presentamos 2 posibles soluciones a la misma pregunta, la primera es dada una consulta que tiene varias palabras buscar toda la consulta y la segunda es hacer varias consultas de cada palabra y ver que documentos son mas frecuentes. Con el segundo método obtenemos mejores resultados y aunque se puede considerar algo “tramposo”.

Los segundos datos los mostramos solo para 1K y [P@5](#) como ejemplo, si se quiere probar se encuentra en el archivo TextSercher2.java

Para el archivo 1K:

P@5

Consulta 1: obama family tree

clueweb09-en0010-79-2218.html 0.7363380789756775
clueweb09-en0001-02-21241.html 0.6691653728485107
clueweb09-en0010-93-11761.html 0.6094216108322144
clueweb09-enwp01-59-16163.html 0.5670822858810425
clueweb09-enwp00-39-9864.html 0.5665059089660645

consulta: obama family tree

clueweb09-en0010-79-2218.html 2.2090142369270325
clueweb09-en0001-02-21241.html 2.0074961185455322
clueweb09-en0010-93-11761.html 1.828264832496643
clueweb09-enwp01-59-16163.html 1.7012468576431274
clueweb09-enwp00-39-9864.html 1.6995177268981934

Consulta 2: french lick resort and casino

clueweb09-en0006-85-33176.html 1.1752631664276123
clueweb09-en0004-03-39951.html 1.0292868614196777
clueweb09-en0005-00-22740.html 1.0241541862487793
clueweb09-en0000-99-14575.html 0.9189548492431641
clueweb09-en0000-37-9377.html 0.9155914187431335

consulta: french lick resort and casino

clueweb09-en0006-85-33176.html 5.8763158321380615
clueweb09-en0004-03-39951.html 5.146434307098389
clueweb09-en0005-00-22740.html 5.1207709312438965
clueweb09-en0000-99-14575.html 4.59477424621582
clueweb09-en0000-37-9377.html 4.577957093715668

Consulta 3: getting organized

clueweb09-en0002-44-10207.html 0.9979654550552368
clueweb09-en0008-14-3912.html 0.7825719714164734
clueweb09-en0002-44-10184.html 0.781821608543396
clueweb09-en0002-44-10182.html 0.781821608543396
clueweb09-en0002-44-10183.html 0.781821608543396

consulta: getting organized

clueweb09-en0002-44-10207.html 1.9959309101104736
clueweb09-en0008-14-3912.html 1.5651439428329468
clueweb09-en0002-44-10184.html 1.563643217086792
clueweb09-en0002-44-10182.html 1.563643217086792
clueweb09-en0002-44-10183.html 1.563643217086792

Consulta 4: toilet

clueweb09-en0008-76-36759.html 1.2838401794433594
clueweb09-en0001-14-26411.html 1.142032504081726
clueweb09-en0000-22-28298.html 1.1414642333984375

clueweb09-en0010-46-16451.html 1.1272783279418945
clueweb09-en0005-22-25429.html 0.8793469667434692

consulta: toilet

clueweb09-en0008-76-36759.html 1.2838401794433594
clueweb09-en0001-14-26411.html 1.142032504081726
clueweb09-en0000-22-28298.html 1.1414642333984375
clueweb09-en0010-46-16451.html 1.1272783279418945
clueweb09-en0005-22-25429.html 0.8793469667434692

Consulta 5: mitchell college

clueweb09-en0007-60-34059.html 0.9017983675003052
clueweb09-en0004-19-8697.html 0.8951672911643982
clueweb09-en0003-77-29412.html 0.8370717763900757
clueweb09-enwp02-27-3432.html 0.761343240737915
clueweb09-enwp01-25-3432.html 0.761343240737915

consulta: mitchell college

clueweb09-en0007-60-34059.html 1.8035967350006104
clueweb09-en0004-19-8697.html 1.7903345823287964
clueweb09-en0003-77-29412.html 1.6741435527801514
clueweb09-enwp02-27-3432.html 1.52268648147583
clueweb09-enwp01-25-3432.html 1.52268648147583

P@10

Consulta 1: obama family tree

clueweb09-en0010-79-2218.html 0.7363380789756775
clueweb09-en0001-02-21241.html 0.6691653728485107
clueweb09-en0010-93-11761.html 0.6094216108322144
clueweb09-enwp01-59-16163.html 0.5670822858810425
clueweb09-enwp00-39-9864.html 0.5665059089660645
clueweb09-enwp03-00-6901.html 0.5665059089660645
clueweb09-enwp00-81-12487.html 0.5665059089660645
clueweb09-enwp00-04-9625.html 0.5665059089660645
clueweb09-enwp00-00-9498.html 0.5665059089660645
clueweb09-enwp01-75-20596.html 0.5665059089660645

Consulta 2: french lick resort and casino

clueweb09-en0006-85-33176.html 1.1752631664276123
clueweb09-en0004-03-39951.html 1.0292868614196777
clueweb09-en0005-00-22740.html 1.0241541862487793
clueweb09-en0000-99-14575.html 0.9189548492431641
clueweb09-en0000-37-9377.html 0.9155914187431335
clueweb09-en0006-85-33174.html 0.8984381556510925
clueweb09-en0000-99-778.html 0.8624207973480225
clueweb09-en0003-75-16320.html 0.8502689599990845
clueweb09-en0000-97-28663.html 0.8200163245201111
clueweb09-en0000-97-28664.html 0.8200163245201111

Consulta 3: getting organized

clueweb09-en0002-44-10207.html 0.9979654550552368
clueweb09-en0008-14-3912.html 0.7825719714164734
clueweb09-en0002-44-10184.html 0.781821608543396
clueweb09-en0002-44-10182.html 0.781821608543396
clueweb09-en0002-44-10183.html 0.781821608543396
clueweb09-en0010-68-5220.html 0.7361009120941162
clueweb09-en0002-90-13719.html 0.6983911395072937

clueweb09-en0011-41-711.html 0.6361598372459412
clueweb09-en0006-56-29452.html 0.5891280174255371
clueweb09-en0003-35-33123.html 0.5834686756134033

Consulta 4: toilet

clueweb09-en0008-76-36759.html 1.2838401794433594
clueweb09-en0001-14-26411.html 1.142032504081726
clueweb09-en0000-22-28298.html 1.1414642333984375
clueweb09-en0010-46-16451.html 1.1272783279418945
clueweb09-en0005-22-25429.html 0.8793469667434692
clueweb09-en0005-02-11979.html 0.8703277707099915
clueweb09-en0008-14-32550.html 0.8471148014068604
clueweb09-en0000-40-14160.html 0.7818516492843628
clueweb09-en0011-22-13728.html 0.7775908708572388
clueweb09-en0000-08-10770.html 0.7684581279754639

Consulta 5: mitchell college

clueweb09-en0007-60-34059.html 0.9017983675003052
clueweb09-en0004-19-8697.html 0.8951672911643982
clueweb09-en0003-77-29412.html 0.8370717763900757
clueweb09-enwp02-27-3432.html 0.761343240737915
clueweb09-enwp01-25-3432.html 0.761343240737915
clueweb09-en0005-11-41513.html 0.7530998587608337
clueweb09-en0006-09-11849.html 0.7373160719871521
clueweb09-en0009-23-30659.html 0.7198832035064697
clueweb09-en0010-43-1709.html 0.7184184789657593
clueweb09-en0006-09-11869.html 0.7143234014511108

Para el de 10K:

P@5

Consulta 1: obama family tree

clueweb09-en0004-41-04748.html 0.7620208263397217
clueweb09-en0000-80-03184.html 0.5629175305366516
clueweb09-en0003-64-15490.html 0.5526552200317383
clueweb09-en0003-64-15476.html 0.5500283241271973
clueweb09-en0006-52-11056.html 0.5064629912376404

Consulta 2: french lick resort and casino

clueweb09-en0003-75-16320.html 1.1570760011672974
clueweb09-en0000-99-00778.html 1.0005370378494263
clueweb09-en0000-99-00782.html 0.9412717223167419
clueweb09-en0000-37-09377.html 0.8818815350532532
clueweb09-en0000-99-14575.html 0.8816102743148804

Consulta 3: getting organized

clueweb09-en0007-04-05869.html 0.3647094964981079
clueweb09-en0010-26-19905.html 0.23209306597709656
clueweb09-en0001-45-08853.html 0.1547287106513977
clueweb09-en0007-72-12548.html 0.0997980386018753
clueweb09-en0008-89-29328.html 0.09336545318365097

Consulta 4: toilet

clueweb09-en0001-13-07399.html 1.4198493957519531
clueweb09-en0004-90-07844.html 1.4177566766738892
clueweb09-en0001-14-26411.html 1.417407751083374
clueweb09-en0000-22-28298.html 1.4170585870742798
clueweb09-en0008-76-36759.html 1.3298248052597046

Consulta 5: mitchell college

clueweb09-en0004-13-23957.html 0.663654625415802
clueweb09-en0003-64-22468.html 0.5229780673980713
clueweb09-en0009-52-04965.html 0.504435122013092

clueweb09-en0009-52-04612.html 0.40372323989868164
clueweb09-en0010-16-23037.html 0.3250654637813568
 Consulta 6: kcs
 clueweb09-en0008-72-25615.html 0.7065712213516235
clueweb09-en0008-72-25608.html 0.636894702911377
clueweb09-en0003-96-19588.html 0.5888093709945679
clueweb09-en0008-72-25613.html 0.5769129991531372
clueweb09-en0008-97-20355.html 0.5548961758613586
 Consulta 7: air travel information
 clueweb09-en0001-18-32624.html 0.41289520263671875
clueweb09-en0001-66-18805.html 0.3974624574184418
clueweb09-en0001-66-18806.html 0.39633917808532715
clueweb09-en0001-66-18809.html 0.39370590448379517
clueweb09-en0005-04-03355.html 0.3865308165550232
 Consulta 8: appraisals
 clueweb09-en0010-58-03630.html 0.3538493514060974
clueweb09-en0005-27-13117.html 0.31970566511154175
clueweb09-en0008-73-33466.html 0.3164924681186676
clueweb09-en0004-17-29299.html 0.3032994568347931
clueweb09-en0007-99-25654.html 0.299057275056839
 Consulta 9: used car parts
 clueweb09-en0000-55-00481.html 0.6461268663406372
clueweb09-en0000-07-00284.html 0.5947952270507812
clueweb09-en0004-74-31016.html 0.5669481158256531
clueweb09-en0003-07-27524.html 0.5667840242385864
clueweb09-en0007-59-38033.html 0.545210063457489
 Consulta 10: cheap internet
 clueweb09-en0001-18-32681.html 1.1102097034454346
clueweb09-en0003-07-27266.html 1.0785720348358154
clueweb09-en0000-71-06224.html 0.9634552001953125
clueweb09-en0004-41-04645.html 0.9505078196525574
clueweb09-en0004-28-00281.html 0.9476161599159241

P@10

Consulta 1:

 clueweb09-en0004-41-04748.html 0.7620208263397217
clueweb09-en0000-80-03184.html 0.5629175305366516
clueweb09-en0003-64-15490.html 0.5526552200317383
clueweb09-en0003-64-15476.html 0.5500283241271973
clueweb09-en0006-52-11056.html 0.5064629912376404
clueweb09-en0008-45-29117.html 0.4407825767993927
clueweb09-en0004-65-06301.html 0.37335431575775146
clueweb09-en0009-67-16390.html 0.34857022762298584
clueweb09-en0003-35-18580.html 0.30844438076019287
clueweb09-en0000-39-00003.html 0.3071419894695282

Consulta 2:

 clueweb09-en0003-75-16320.html 1.1570760011672974
clueweb09-en0000-99-00778.html 1.0005370378494263
clueweb09-en0000-99-00782.html 0.9412717223167419
clueweb09-en0000-37-09377.html 0.8818815350532532
clueweb09-en0000-99-14575.html 0.8816102743148804
clueweb09-en0004-65-06602.html 0.8779970407485962
clueweb09-en0001-65-10393.html 0.8639882802963257
clueweb09-en0001-82-16970.html 0.8518757820129395
clueweb09-en0003-85-29600.html 0.8470951914787292
clueweb09-en0008-73-34552.html 0.7629407048225403

Consulta 3:

 clueweb09-en0007-04-05869.html 0.3647094964981079
clueweb09-en0010-26-19905.html 0.23209306597709656
clueweb09-en0001-45-08853.html 0.1547287106513977
clueweb09-en0007-72-12548.html 0.0997980386018753
clueweb09-en0008-89-29328.html 0.09336545318365097

clueweb09-en0009-81-19920.html 0.08252418786287308
clueweb09-en0004-01-12064.html 0.08085685968399048
clueweb09-en0005-99-18498.html 0.08085685968399048
clueweb09-en0004-48-14195.html 0.08085685968399048
clueweb09-en0004-75-26250.html 0.08085685968399048

Consulta 4:

clueweb09-en0001-13-07399.html 1.4198493957519531
clueweb09-en0004-90-07844.html 1.4177566766738892
clueweb09-en0001-14-26411.html 1.417407751083374
clueweb09-en0000-22-28298.html 1.4170585870742798
clueweb09-en0008-76-36759.html 1.3298248052597046
clueweb09-en0004-41-12172.html 0.9196404218673706
clueweb09-en0011-07-01195.html 0.7988637685775757
clueweb09-en0006-45-24128.html 0.7679122686386108
clueweb09-en0003-27-18446.html 0.7666739821434021
clueweb09-en0011-82-35253.html 0.7397647500038147

Consulta 5:

clueweb09-en0004-13-23957.html 0.663654625415802
clueweb09-en0003-64-22468.html 0.5229780673980713
clueweb09-en0009-52-04965.html 0.504435122013092
clueweb09-en0009-52-04612.html 0.40372323989868164
clueweb09-en0010-16-23037.html 0.3250654637813568
clueweb09-en0001-90-18773.html 0.31171685457229614
clueweb09-en0006-97-08104.html 0.25854426622390747
clueweb09-en0005-79-00135.html 0.22775287926197052
clueweb09-en0005-06-38180.html 0.2241920679807663
clueweb09-en0004-18-29141.html 0.20455938577651978

Consulta 6:

clueweb09-en0008-72-25615.html 0.7065712213516235
clueweb09-en0008-72-25608.html 0.636894702911377
clueweb09-en0003-96-19588.html 0.5888093709945679
clueweb09-en0008-72-25613.html 0.5769129991531372
clueweb09-en0008-97-20355.html 0.5548961758613586
clueweb09-en0008-72-25610.html 0.5396527051925659
clueweb09-en0011-27-10876.html 0.5307455658912659
clueweb09-en0011-83-32289.html 0.5266470909118652
clueweb09-en0008-62-31894.html 0.49417033791542053
clueweb09-en0008-72-25612.html 0.4882149398326874

Consulta 7:

clueweb09-en0001-18-32624.html 0.41289520263671875
clueweb09-en0001-66-18805.html 0.3974624574184418
clueweb09-en0001-66-18806.html 0.39633917808532715
clueweb09-en0001-66-18809.html 0.39370590448379517
clueweb09-en0005-04-03355.html 0.3865308165550232
clueweb09-en0009-07-07584.html 0.3845005929470062
clueweb09-en0005-04-03359.html 0.3841155469417572
clueweb09-en0005-04-03354.html 0.3841155469417572
clueweb09-en0001-94-20393.html 0.38084709644317627
clueweb09-en0005-08-32423.html 0.36624258756637573

Consulta 8:

clueweb09-en0010-58-03630.html 0.3538493514060974
clueweb09-en0005-27-13117.html 0.31970566511154175
clueweb09-en0008-73-33466.html 0.3164924681186676
clueweb09-en0004-17-29299.html 0.3032994568347931
clueweb09-en0007-99-25654.html 0.299057275056839
clueweb09-en0009-23-25797.html 0.2712792754173279
clueweb09-en0009-23-25800.html 0.2712792754173279
clueweb09-en0000-21-18630.html 0.2712792754173279
clueweb09-en0005-98-13970.html 0.25576454401016235
clueweb09-en0008-23-30823.html 0.25576454401016235

Consulta 9:

clueweb09-en0000-07-00284.html 0.5947952270507812

clueweb09-en0004-74-31016.html 0.5669481158256531
clueweb09-en0003-07-27524.html 0.5667840242385864
clueweb09-en0007-59-38033.html 0.545210063457489
clueweb09-en0005-71-07801.html 0.5343656539916992
clueweb09-en0000-61-26460.html 0.524873673915863
clueweb09-en0008-00-27021.html 0.5070152282714844
clueweb09-en0008-43-19366.html 0.50069260597229
clueweb09-en0000-21-19692.html 0.483451783657074

Consulta 10:

clueweb09-en0001-18-32681.html 1.1102097034454346
clueweb09-en0003-07-27266.html 1.0785720348358154
clueweb09-en0000-71-06224.html 0.9634552001953125
clueweb09-en0004-41-04645.html 0.9505078196525574
clueweb09-en0004-28-00281.html 0.9476161599159241
clueweb09-en0004-77-15242.html 0.9422887563705444
clueweb09-en0000-95-28647.html 0.8368417024612427
clueweb09-en0001-43-24713.html 0.8326862454414368
clueweb09-en0000-95-28660.html 0.8312958478927612
clueweb09-en0000-95-28646.html 0.8312958478927612