

Title: Laboratory 1

Date: 26/10/21

Student names: Vincent Achukwu, Alan Hian Wu

Examining Document Collection Statistics

Upon examining the information retrieved via the IRModelGenerator, we noticed how when entering a term that is listed in the *Num top terms* section, it would display the document and collection frequency of that term. Certain terms (like Bitcoin) display 0 for the document and collection frequency columns. We believe certain terms won't work since the lab sheet states "The document collection consists of about 500,000 news articles taken mainly from the LA Times newspaper. The edition of the paper containing these news stories was published more than 20 years ago", hence Bitcoin wasn't around then.

We noticed that if you pass a value (e.g 10) to the first field and click *Get term stats*, it would simply display the first 10 of the *Num top terms*. What these values suggest to us for each term is that for each term, the document frequency suggests how many documents contain the term, and the collection frequency is the number of times the term appears within the document. The IRModelGenerator lists these terms ranked from highest to lowest document frequency, as well as being able to filter how many terms you would like to list (i.e 20, 50, 100, 1000). As stated before, you can specify how many terms you want to see the statistics for by specifying the value in the first field after which you can click *Get term stats*.

These values (document and collection frequency) can contribute to retrieving relevant documents by observing how certain terms appear in certain sites. For example, if we want to travel and search "London", according to the IRModelGenerator, this term has a document frequency of 213197, and a collection frequency of 298164.

Interactive Searching using Lucene

Upon entering a search term and specifying the number of documents to retrieve, We could select an IR Model for the retrieval method. We noticed how the search term was highlighted in **bold** in the snippets. This is a summarization technique to allow users to see how their search term relates to the document they may be searching for. For example, when we search "time", it displays the relevant documents associated with the search term, highlighting the appearance of the search term in the snippets in **bold**, and the search term is also present within the original text of the document after clicking on the document ID. Within the term frequency vector, it displayed the words that were present in the IR Model Generator and the retrieved document with their document and collection frequencies.

BM25

Using BM25, we were able to change the parameters (k and b). After entering a search term, it shows part of the document text. To read more of it, we had to click on the document ID.

As stated in the notes, the b parameter varies between the range 0 to 1. When using this parameter, we should ask ourselves “When do we think a document will be very long?”. We noticed that the closer to 0 it is, the longer the term frequency vector, and the closer to 1 it is the shorter the term frequency vector. The b is set within the range 0 to 1 to allow for optimal retrieval effectiveness for a document collection using a set of training queries and their corresponding relevance assessments. For the k parameter, we should ask ourselves “when do we think a search term is going to be saturated”. For example “dog” and “dogs” can appear many times, but you may want to have k set to a high value when the text is longer and more diverse. On the other hand, having k lower means it’s very unlikely that a collection of short documents would have “dogs” many times without really being related to the subject of dogs.

TF-IDF

As with the BM25 model, it showed a collection of documents with the snippets available, as well as the term frequency vectors and being able to view the full original text of the snippets. Clicking on the document ID allowed us to see the full text of the snippets.

Main stuff to cover

- Answer: **What do the values suggest to you** for “View Collection Stats”

And the following **Can you make any observations about their distributions in the document collection and about their likely usefulness for information retrieval**
how you might expect them to contribute to retrieving relevant documents.

- Answer: **See if you can gauge how the term weights and document lengths might have produced the ranked output that your see**

Try entering more search requests to explore this relationship further, see if the ranking behaviour for these new queries is consistent with your observations relating to the results for the previous queries.

Change to BM25, what changed?

Explain the changes in terms of k and b , may need many tries.

- **Again enter your observations about the relationship between the search requests, the snippets, and the documents into your laboratory report, including examples, of queries, doc IDs and their rankings, and snippets to illustrate your observations.**

Section 4

Looking at the View Collection Stats, time and page are the most frequently used words, which makes sense, since these articles are all mainly taken from the “LA TIMES”, so time will appear at least once, and it also helps that the word “time” is the 68th most used word in the English language according to espressoenglish.net, the word “page”, “year” and “report” all having a high frequency all makes sense since these are commonly used words in newspapers.

Term	Doc Freq	Coll Freq	
time	360476	684393	
page	347260	383440	
year	266010	827942	61. some
financi	247956	344316	
countri	229732	417625	62. her
report	227965	428085	
94	227231	406769	63. would
part	224425	354664	
london	213197	298164	
ft	211966	242930	64. make
_an	210204	210215	
1	202111	499593	65. like
industri	199735	382327	
state	190898	550368	66. him
unit	181318	323331	
1994	173764	272099	
2	171375	404964	67. into
type	164542	186382	
compani	158337	525676	68. time
govern	158244	449893	

Some more interesting words appear as well, words like “financi” which is not a proper word by itself but is part of words like financial, this at first lead me to think that there is a word limit of 7, where the returned results are truncated, but results like the follow disproves that believe.

issu	124264	251727
document	89818	136806

The first conclusion is that these half words appear more often than the full words, to test this theory we can just simply look the full words up and compare them to the half words.

Enter a term
Doc freq: 247956 Coll Freq: 344316

financi 247956 344316

Enter a term
Doc freq: 124264 Coll Freq: 251727

issu 124264 251727

The result is that the half words and full words return the same result,
So the next theory is that there is some optimization going on here, where words like “financi” and “financial” is grouped together as the same thing, this theory is proven when looking up the words “document”, “documents”, “document’s”, “documented” which is basically the same thing.

Enter a term
Doc freq: 89818 Coll Freq: 136806

Enter a term
Doc freq: 89818 Coll Freq: 136806

Enter a term
Doc freq: 89818 Coll Freq: 136806

Enter a term
Doc freq: 89818 Coll Freq: 136806

The distribution of these words in all the documents are heavily influenced by the fact that these are newspapers, so big news events or ongoing news events will be favoured since there is more of them; this seems to hold true with the word “1994” and “94” being questionably frequent in the collection, until you consider there might be a significant amount of news articles from the year 1994.

94 227231 406769
1994 173764 272099

From looking at all the statistics from the collection itself we can conclude that, searching anything later than the year 2000 will probably return no results since the base of the collection will not have news articles from this time period, searching big news events around the 1994 time period will return a lot of results, searching up words related to “time”, “page” or other popular terms might not return the result you want, since nearly every newspaper in the collection will contain these terms so for the search algorithm these terms are very vague.

Section 5

Since this collection of articles are from 20 years ago and is based in the USA We decided to look up some big events that took place during that time period and area. Looking up on google and then that led to <https://www.infoplease.com/year/1994> which shows us the biggest news event in the most popular year in our collection.

U.S. Events

- Olympic figure skater [Nancy Kerrigan](#) attacked (Jan. 6); three arrested in attack (Jan. 13).
- Aldrich Ames, high [C.I.A.](#) official, charged with [spying](#) for Soviets (Feb. 22).
- Four convicted in [World Trade Center bombing](#) (March 4).
- [Clinton](#) accused of [sexual harassment](#) while Governor of [Arkansas](#) (May 6).
- [O. J. Simpson](#) arrested in killings of wife, Nicole Brown Simpson, and friend, Ronald Goldman (June 18).
- [US Supreme Court](#) approves limit on abortion protests (June 30).
- Senate confirms [Stephen G. Breyer](#) for Supreme Court (July 29).
- Major league [baseball players strike](#) (Aug. 13).
- Newt Gingrich named [House Speaker](#) (Dec. 5).

U.S. Statistics

President: [William J. Clinton](#)

Vice President: [Albert Gore, Jr.](#)

Population: 260,289,237

Life expectancy: 75.7 years

Violent Crime Rate (per 1,000): 53.7

Property Crime Rate (per 1,000) 46.6

[More U.S. Statistics...](#)

We decided to look up Clinton on TF-IDF, hoping that the search algorithm will return relevant articles to this event.

[illegible]

Looking at the top 3 results, the first one is about world politics with president Bill Clinton. The second one is about Bill Clinton's foreign policy, and the third one is about Bill Clinton's Whitewater controversy. None of which is the article we were hoping to find, by examining the snippets we see that the top 2 results have 9 occurrences of Clinton in it and the 3rd one only has 6, while astonishingly the 4th one with 13 occurrences is not ranked first, by closely looking at the articles that returned we noticed that the top results are all shorter than the 4th article so the tfidf might have given that article a low ranking score as it is longer than the others.

Using BM-25 $k=1.2$ $b=0.75$ the results for the same query is

[illegible][illegible][illegible][illegible]

Next we wanted to change the values of “k” and “b”, with “k” being the variable that sets the saturation and “b” being the variable that sets the length condition of the articles, for this case we first move the value of “k” up without changing the value of “b”, so in theory the top results should shift in favour of more occurrences of the term “clinton”.

IR Model: #docs to retrieve: BM25 parameters: k b **FT921-12198**

Term Freq Vector:

clinton:16, 45.894596 mr:13, 4.2022786 arkansa:4, 326.0216 campaign:4, 19.610685 tv:4, 36.63164 american:3, 9.836388 appear:3, 12.562856 bill:3, 9.10722 confess:3, 199.30377 democrat:3, 14.2563505 flower:3, 130.89343 marriag:3, 144.30464 ms:3, 58.4695 question:3, 8.762713 respons:3, 7.309497 support:3, 5.7452493 tape:3, 86.88189 time:3, 1.46516 voter:3, 48.277424 week:3, 4.988524 admit:2, 27.425226 affair:2, 8.156456 boston:2, 90.221214 cb:2, 183.77 cent:2, 5.2740607 claim:2, 11.167484 confer:2, 13.39509 employe:2, 20.076595 fell:2, 16.969381 gamb:2, 150.81525 governor:2, 35.86304 hampshir:2, 363.49277 hillari:2, 1650.4844 hotel:2, 33.35786 interview:2, 18.791538 local:2, 8.321071 ly:2, 190.73854 media:2, 25.853199 minut:2, 20.530807 nomin:2, 64.66148 poll:2, 38.48965 press:2, 9.285752 primari:2, 41.028122 report:2, 2.316825 reuter:2, 42.908035 rock:2, 54.634842 tonight:2, 83.70127 wife:2, 38.24162 year:2, 1.9854704 yesterdai:2, 5.5204186 _an:1, 2.512583 absolut:1, 47.35542 accus:1, 25.814028 acknowledg:1, 35.499058 adulteri:1, 3220.4573 advanc:1, 15.153789 advis:1, 21.965273 ago:1, 8.942231 alleg:1, 27.432348 anger:1, 104.19314 appeal:1, 24.501532 attract:1, 19.151576 avoid:1, 21.536251 barber:1, 311.41214 begin:1, 8.61127 cabaret:1, 1748.8577 calm:1, 97.734085 caus:1, 19.695518 clear:1, 10.987892 close:1, 6.3912654 comfort:1, 53.77813 commit:1, 15.753124 conduct:1, 15.439517 convers:1, 41.817497 coupl:1, 33.586964 credibl:1, 77.76134 critic:1, 13.606982 cut:1, 9.034468 dai:1, 4.1854615 daili:1, 6.8684325 data:1, 15.732477 deni:1, 22.6822 disappoint:1, 41.980366 discuss:1, 8.72133 draw:1, 26.917843 drown:1, 406.89908 earli:1, 9.261815 elect:1, 10.535707 equip:1, 15.38958 excelsior:1, 2296.3262 excerpt:1, 40.64607 extraordinari:1, 64.15098 fair:1, 29.80054 fall:1, 9.873163 februari:1, 10.864704 financi:1, 2.1300352 find:1, 8.781362 friend:1, 26.28683 front:1, 18.425083 ft:1, 2.4916968 game:1, 17.419935 gener:1, 3.3491335 gennif:1, 44012.918 good:1, 5.906387 gotcha:1, 17605.166 group:1, 4.1013465 he'd:1, 254.16507 held:1, 8.504089 hope:1, 8.490556 huge:1, 35.802265 humour:1, 629.5054 inconclus:1, 994.6422 jan:1, 11.734431 joint:1, 15.755004 judg:1, 25.404282 kark:1, 132038.75 latest:1, 19.349173 li:1, 39.344086 light:1, 18.296785 lionel:1, 414.89 london:1, 2.4773097 long:1, 5.58575 lover:1, 213.48221 maintain:1, 12.542866 mari:1, 93.36309 massachusett:1, 207.28218 messag:1, 36.540405 met:1, 18.802242 monei:1, 9.500558 month:1, 4.334291 new:1, 3.8393402 night:1, 12.521456 offer:1, 7.7248397 omit:1, 31.469643 open:1, 6.1597447 page:1, 1.5209209 pain:1, 57.408154 part:1, 2.3533697 parti:1, 6.676801 paul:1, 29.962841 peopl:1, 3.824605 percentag:1, 36.575832 photograph:1, 66.417885 pin:1, 177.88986 place:1, 6.585556 plai:1, 10.0333395 point:1, 5.6887507 politician:1, 34.269077 popular:1, 26.004677 prepar:1, 10.330256 presidenti:1, 35.287968 prime:1, 12.409074 privaci:1, 241.0566 program:1, 21.386257 protect:1, 10.944857 prove:1, 20.4624 public:1, 4.7937393 publish:1, 11.970604 put:1, 7.017552 realist:1, 87.61695 record:1, 8.502173 refus:1, 20.485416 republican:1, 65.123924 research:1, 15.522557 respect:1, 16.096888 reveal:1, 30.543314 risk:1, 18.055347 rocki:1, 294.07294 rumour:1, 134.93996 runner:1, 145.25716 salvag:1, 367.02917 scandal:1, 85.103935 scrutini:1, 214.34862 senat:1, 49.99574 sex:1, 113.046875 sharpli:1, 38.706852 singer:1, 119.95344 sixti:1, 340.52548 sold:1, 19.257456 state:1, 2.766687 station:1, 20.627026 stori:1, 17.299541 supermarket:1, 150.51439 surfac:1, 47.95306 survei:1, 21.752678 tabloid:1, 512.2745 telephon:1, 19.682306 televis:1, 15.581632 thing:1, 10.020966 thought:1, 17.126759 tsonga:1, 5127.7183 turn:1, 8.733588 urg:1, 24.178493 viewer:1, 187.22261 washington:1, 11.138751 work:1, 4.2602763 wrongdo:1, 464.51627 york:1, 14.04929

FT921-12198 AN-CA1BTAD7FT 920128 FT 28 JAN 92 / Clinton pins campaign hopes on TV confession By LIONEL BARBER and REUTER WASHINGTON, BOSTON GOVERNOR Bill Clinton of Arkansas was gambling yesterday, Hillary, Mr Clinton admitted his marriage had been rocky, but he denied allegations of a 12-year, the employee, last night held a news conference in New York in which she accused Mr Clinton of lying when he claimed they were just friends. She said: 'I was Bill Clinton's lover for 12 years'

FBIS3-59184

Term Freq Vector:

clinton:13, 45.894596 presid:8, 5.118029 plan:7, 4.645122 state:7, 2.766687 popular:6, 26.004677 republican:5, 65.123924 american:4, 9.836388 unit:4, 2.9128659 foreign:3, 5.7644367 reform:3, 15.490687 speech:3, 34.701378 union:3, 10.360044 address:2, 14.078502 bill:2, 9.10722 brussel:2, 48.81736 compromis:2, 53.756233 crime:2, 34.630844 de:2, 16.868572 domest:2, 14.97038 dutch:2, 78.81734 fight:2, 18.80492 jan:2, 11.734431 law:2, 8.890151 morgan:2, 7.766.9854 opposit:2, 18.34636 privat:2, 10.977386 support:2, 5.7452493 talk:2, 7.774646 theme:2, 75.67775 wing:2, 49.452717 absolut:1, 47.35542 affair:1, 8.156456 anticrim:1, 7882.9106 aris:1, 62.128574 articl:1, 6.381227 averag:1, 13.976052 belgium:1, 85.51733 benefit:1, 12.841426 benelux:1, 546.7443 billion:1, 23.914648 boost:1, 29.499273 brillanti:1, 772.15643 build:1, 8.077741 bush:1, 47.37666 call:1, 5.006873 camp:1, 53.001003 care:1, 19.872635 chapter:1, 54.398495 coast:1, 38.30819 commentari:1, 109.01032 comprehens:1, 46.884598 consider:1, 15.727793 convers:1, 41.817497 cost:1, 6.4599004 daili:1, 6.8684325 danger:1, 24.248428 death:1, 24.477684 demagog:1, 3000.8806 democrat:1, 14.2563505 document:1, 5.88028 doubt:1, 23.915731 east:1, 9.873901 expand:1, 20.779596 face:1, 9.155687 fanat:1, 631.0096 fbi:1, 4.515535 featur:1, 29.954344 financ:1, 8.422719 find:1, 8.781362 floor:1, 41.066402 found:1, 9.5647335 georg:1, 26.284214 good:1, 5.906387 ground:1, 17.080786 health:1, 15.862416 hobbyhors:1, 40627.31 hunt:1, 88.2612 imag:1, 40.36031 inclin:1, 169.82475 includ:1, 3.5955327 irresist:1, 666.86237 irrevoc:1, 747.0368 isol:1, 67.4011 issu:1, 4.2502656 languag:1, 7.2994957 larg:1, 7.562357 leadership:1, 26.693369 left:1, 9.204353 make:1, 3.423398 manag:1, 5.0954638 million:1, 9.764555 monei:1, 9.500558 natur:1, 12.473549 oblig:1, 34.074516 order:1, 6.43401 origin:1, 16.331324 overtak:1, 782.45184 pass:1, 14.946232 peopl:1, 3.824605 plai:1, 10.0333395 policemen:1, 231.24124 polici:1, 6.303694 poverty:1, 110.8405 predecessor:1, 124.85934 priorit:1, 31.793583 prison:1, 46.484333 provid:1, 5.7699156 question:1, 8.762713 quicker:1, 433.6248 rais:1, 9.425112 reach:1, 9.473123 real:1, 11.970332 recent:1, 5.781474 remain:1, 6.573589 report:1, 2.316825 review:1, 10.82773 rival:1, 39.966328 scarc:1, 160.09549 sector:1, 11.133116 secur:1, 6.0898566 senat:1, 49.99574 sentenc:1, 47.491684 short:1, 12.639521 shown:1, 32.956135 shut:1, 89.320984 significanti:1, 49.217686 skilfulli:1, 2824.3584 social:1, 11.985272 splendid:1, 358.31412 stake:1, 26.012362 stricken:1, 418.50635 strike:1, 27.94471 strikingli:1, 1408.4133 subject:1, 12.701835 system:1, 6.752001 text:1, 4.375694 time:1, 1.46516 topic:1, 73.192215 tradit:1, 22.989248 train:1, 16.756187 turn:1, 8.733588 type.bfn:1, 9.673877 type.daili:1, 10.209642 u.:1, 9.52008 unattribut:1, 137.86348 unclear:1, 139.5021 underlin:1, 98.224846 undoubtedli:1, 142.55196 vagu:1, 213.56854 water:1, 16.030443 west:1, 10.27459 weu:1, 41.757984 work:1, 4.2602763 world:1, 5.1912227 year:1, 1.985470

Belgium Daily Reviews Clinton's State of Union Speech BR2701155994 Brussels DE MORGEN in Dutch 27 Jan 94: "Clinton Skilfully Plays on Popular Themes" [Text] With a brilliantly managed first State of the Union address, U.S. President Bill Clinton has given his image and popularity a large boost. For some time isolation, when talking about the American east and west coasts. Clinton's predecessor, George Bush, was the "foreign affairs" president. With his State of the Union address, Clinton has left

FBIS3-39803

Term Freq Vector:

clinton:9, 45.894596 ciller:8, 918.53046 bosnia:5, 54.376095 minist:5, 6.0379205 prime:5, 12.409074 convers:4, 41.817497 herzegovina:4, 110.77077 telephon:4, 19.682306 ankara:3, 226.28749 develop:3, 4.813267 sarajevo:3, 117.9444 statement:3, 11.338665 u.:3, 9.52008 bill:2, 9.10722 discuss:2, 8.72133 feb:2, 12.911747 letter:2, 14.380173 mr:2, 4.2022786 network:2, 12.110869 presid:2, 5.118029 report:2, 2.316825 tansu:2, 1025.5437 televis:2, 15.581632 trt:2, 862.99835 turkish:2, 114.34402 accompani:1, 38.61347 ad:1, 7.020724 admir:1, 85.53117 afternoon:1, 33.87129 ambassador:1, 55.612823 announc:1, 7.383272 approxim:1, 44.83489 articl:1, 6.381227 attack:1, 17.65165 bhutto:1, 762.1284 center:1, 12.6289425 courag:1, 137.07631 decisi:1, 7.7518234 departur:1, 71.81874 document:1, 5.88028 express:1, 12.860813 fbi:1, 4.515535 februari:1, 10.864704 feel:1, 14.393497 fulli:1, 26.123009 gmt:1, 6.22494 half:1, 9.219939 hope:1, 8.490556 hour:1, 11.597352 inhuman:1, 818.845 innoc:1, 113.606155 issu:1, 4.2502656 languag:1, 7.2994957 last:1, 74.56657 launch:1, 18.512917 meet:1, 4.9425874 minut:1, 20.530807 nato:1, 70.233376 pakistani:1, 350.4678 palac:1, 116.46196 peopl:1, 3.824605 place:1, 6.585556 presidenti:1, 35.287968 press:1, 9.285752 read:1, 17.73999 regard:1, 23.377966 relai:1, 207.77144 share:1, 6.0378513 state:1, 2.766687 still:1, 1666.1041 stress:1, 20.094164 struggl:1, 26.643545 subject:1, 12.701835 terror:1, 76.6553 text:1, 4.375694 thank:1, 299.57742 turkei:1, 75.69944 type.bfn:1, 9.673877 type.daili:1, 10.209642 visit:1, 12.670449 weu:1, 41.75798

Clinton Telephones Ciller To Discuss Bosnia TA1802194494 Ankara TRT Television Network in Turkish 1800 GMT-read report over stills of Prime Minister Tansu Ciller and U.S. President Bill Clinton [Text] U.S. President Bill Clinton telephoned Prime Minister Tansu Ciller this afternoon and conversed after her conversation with Clinton, Ciller said: Both Mr. Clinton and I are hopeful about developments in Bosnia-Herzegovina following the NATO decision on the subject. The statement added: Mr. Clinton

The result is that nothing changed, this might be because the length of these articles are just too short, since setting the value of k to a higher value is generally used for much longer texts. So the opposite of this situation should be that if we set k to none we should get results that has no occurrence of the term "Clinton"



Clinton

IR Model: #docs to retrieve:

BM25 parameters: k b

• [FT921-4087](#)

Term Freq Vector:

mr:14, 4.2022786 cent:11, 5.2740607 bush:8, 47.37666 presid:6, 5.118029 poll:5, 38.48965 democrat:4, 14.2563505 nixon:4, 388.34927 clinton:3, 45.894596 found:3, 9.5647335 lost:3, 14.927645 previou:3, 16.443182 time:3, 1.46516 tsonga:3, 5127.7183 approv:2, 10.363296 assist:2, 12.435955 drop:2, 16.252424 financi:2, 2.1300352 issu:2, 4.2502656 lose:2, 20.428368 novemb:2, 13.215108 person:2, 7.458447 rate:2, 7.3450756 receiv:2, 6.547349 remind:2, 63.36593 republican:2, 65.123924 russi:2, 17.976685 score:2, 29.272017 super:2, 107.63297 surve:2, 21.752678 tuesdai:2, 10.767905 washington:2, 11.138751 yesterdai:2, 5.5204186 _an:1, 2.512583 abc:1, 226.48157 administr:1, 7.5575953 adult:1, 82.78291 ahead:1, 16.223469 argu:1, 21.430513 arkansa:1, 326.0216 beaten:1, 169.38904 bill:1, 9.10722 blast:1, 122.62712 button:1, 296.8831 campaign:1, 19.610685 carter:1, 172.9954 china:1, 20.070492 comfort:1, 53.778173 comparis:1, 73.202354 conduct:1, 15.439517 conspicu:1, 345.19934 critic:1, 13.606982 defeat:1, 39.491177 devast:1, 156.3514 divis:1, 14.026158 edg:1, 44.690727 editor:1, 23.728771 end:1, 4.1984706 eros:1, 305.46848 error:1, 72.97983 establish:1, 8.158598 find:1, 8.781362 ford:1, 102.23674 foreign:1, 5.7644367 front:1, 18.425083 ft:1, 2.4916968 gener:1, 3.3491335 georg:1, 26.284214 gerald:1, 196.34015 govern:1, 35.86304 greater:1, 17.895674 harri:1, 81.93531 held:1, 8.504089 hot:1, 53.022285 humanitarian:1, 123.31426 identifi:1, 20.215685 inadequ:1, 109.03282 incumb:1, 211.77025 infinit:1, 660.1937 insid:1, 31.629835 intervent:1, 65.79731 invit:1, 26.630112 jimmi:1, 163.31323 johnson:1, 61.801426 jurek:1, 551.8861 lead:1, 7.0955195 london:1, 2.4773097 lowest:1, 67.212395 lyndon:1, 1576.582 main:1, 10.240127 mar:1, 11.091033 march:1, 7.452974 margin:1, 29.121912 martin:1, 43.28075 massachusett:1, 207.28218 minu:1, 255.3941 month:1, 4.334291 nation:1, 3.5148706 new:1, 3.8393402 opinion:1, 18.870766 opposit:1, 18.34636 page:1, 1.5209209 parti:1, 6.676801 pathet:1, 1027.539 paul:1, 29.962841 perform:1, 9.756618 point:1, 5.6887507 polici:1, 6.303694 polit:1, 6.0952682 poor:1, 25.175413 post:1, 13.686317 present:1, 7.5913787 primari:1, 41.028122 public:1, 4.7937393 put:1, 7.017552 quash:1, 1160.7803 question:1, 8.762713 randomli:1, 1216.947 reagan:1, 140.72874 recent:1, 5.781474 record:1, 8.502173 regist:1, 20.461607 relect:1, 40627.31 respons:1, 7.309497 review:1, 10.82773 richard:1, 22.26717 rival:1, 39.966328 ronald:1, 103.19559 runner:1, 145.25716 sampl:1, 82.062614 seek:1, 12.246788 select:1, 21.640375 senat:1, 49.99574 sentiment:1, 71.04587 seri:1, 19.016167 serv:1, 13.570272 sharp:1, 32.124264 show:1, 6.565907 snapshot:1, 1104.9268 soviet:1, 27.978758 spring:1, 31.736269 subject:1, 12.701835 suggest:1, 11.045801 support:1, 5.7452493 take:1, 8.417215 truman:1, 1458.9917 unhelp:1, 2200.6458 union:1, 10.360044 unnam:1, 528.68365 upris:1, 309.4054 victori:1, 25.04885 wider:1, 79.54141 withdrew:1, 203.6064 won:1, 19.278543 wrote:1, 53.90987 yeltsin:1, 70.4394

Clinton of Arkansas and former Senator Paul Tsongas of Massachusetts. The poll also showed Mr Bush's from March 4-8, put him behind Mr Clinton by 46 to 44 per cent and behind Mr Tsongas by a wider political reviews for his campaign performance. But it also indicates that both Mr Clinton and Mr

• [FT921-5900](#)

Term Freq Vector:

mr:9, 4.2022786 maryland:7, 156.72255 tsonga:7, 5127.7183 voter:7, 48.277424 clinton:6, 45.894596 primari:6, 41.028122 state:6, 2.766687 campaign:5, 19.610685 govern:3, 3.3375988 nation:3, 3.5148706 pollit:3, 6.0952682 time:3, 1.46516 ahead:2, 16.223469 apathi:2, 1111.9053 awai:2, 11237.341 base:2, 5.1279674 black:2, 23.078653 busi:2, 4.834238 cent:2, 5.2740607 charact:2, 49.185604 church:2, 65.7318 citi:2, 7.320134 cohen:2, 229.93253 countri:2, 2.299005 curb:2, 79.57737 democrat:2, 14.2563505 give:2, 8.017168 govern:2, 35.86304 hampshir:2, 363.49277 issu:2, 4.2502656 job:2, 11.739648 larg:2, 7.562357 life:2, 10.455823 make:2, 3.423398 parti:2, 6.676801 paul:2, 29.962841 presidenti:2, 35.287968 promis:2, 19.847244 ralli:2, 43.440945 recent:2, 5.781474 respond:2, 22.73003 senat:2, 49.99574 southern:2, 19.842024 support:2, 5.7452493 tuesdai:2, 10.767905 win:2, 15.142058 won:2, 19.278543 worker:2, 17.701344 _an:1, 2.512583 absorb:1, 93.28064 activist:1, 93.49531 affluent:1, 407.527 alleg:1, 27.432348 american:1, 9.836388 amorph:1, 2200.6458 anger:1, 104.19314 appeal:1, 24.501532 attent:1, 20.71196 attract:1, 19.51576 averag:1, 13.976052 awar:1, 33.393715 ballot:1, 90.052 balltimor:1, 229.5328 bascom:1, 132038.75 basic:1, 19.932634 believ:1, 11.696231 beneficiari:1, 169.6611 big:1, 10.837283 bill:1, 9.10722 blue:1, 38.89785 bob:1, 49.64796 brought:1, 18.884935 bush:1, 47.37666 call:1, 5.006873 capita:1, 205.74796 care:1, 19.872635 carri:1, 10.239928 cast:1, 46.199703 cecbhaamft:1, 528155.0 challeng:1, 22.210985 chanc:1, 20.90544 charg:1, 9.539682 charisma:1, 1400.9417 club:1, 35.432377 cold:1, 50.754852 collar:1, 242.27293 commit:1, 15.753124 common:1, 17.252073 commun:1, 7.398994 concern:1, 8.394739 condit:1, 9.777389 conduct:1, 15.439517 corp:1, 24.178493 coven:1, 617.0035 creat:1, 10.400027 credibl:1, 77.76134 cri:1, 316.63968 crime:1, 34.630844 debat:1, 28.493473 demonst:1, 21.67946 devast:1, 156.3514 devot:1, 67.07582 difficult:1, 26.31958 dishearten:1, 2146.9714 distract:1, 287.3531 distrust:1, 430.79526 dodg:1, 471.98837 draft:1, 28.139751 driven:1, 68.43159 drug:1, 25.26332 dunn:1, 307.06686 econom:1, 5.4726553 elect:1, 10.535707 elig:1, 51.67857 employe:1, 20.076595 endless:1, 282.88965 endors:1, 66.24294 energi:1, 20.008144 england:1, 32.7538 ethnic:1, 65.30106 expect:1, 5.123341 fact:1, 10.156046 farmer:1, 65.89582 fervour:1, 2765.2095 financi:1, 2.1300352 fishermen:1, 378.06372 focu:1, 28.433647 forc:1, 6.037989 forthright:1, 1327.0226 ft:1, 2.4916968 georgia:1, 99.99148 germani:1, 16.240929 gloom:1, 325.8205 greatli:1, 89.685005 half:1, 9.219939 haller:1, 7438.8027 hardship:1, 255.3941 high:1, 5.5923147 hold:1, 7.6489882 hope:1, 8.490556 hundr:1, 34.56738 ignor:1, 66.06066 illinoi:1, 178.97493 increasngli:1, 33.416958 influenti:1, 140.8789 insecur:1, 456.4866 inton:1, 1766.4047 japan:1, 16.46779 jeffre:1, 213.65494 keith:1, 162.01074 kerrei:1, 6070.747 kurk:1, 66019.375 labour:1, 24.294159 leadership:1, 26.693369 liber:1, 23.402826 london:1, 2.4773097 loss:1, 10.411714 lousi:1, 1294.4976 low:1, 12.092014 made:1, 3.9773703 man:1, 13.972355 manufactur:1, 14.407114 mar:1, 11.091033 march:1, 7.452974 marion:1, 585.5377 mayor:1, 52.406727 mean:1, 11.060606 mental:1, 87.73339 messag:1, 36.540405 michigan:1, 178.73265 microcosm:1, 2950.5867 minist:1, 6.0379205 minor:1, 24.928257 mobilis:1, 725.4876 mountain:1, 58.886723 mute:1, 429.74368 nanci:1, 145.09753 nasti:1, 348.15756 nativ:1, 85.076515 nebraska:1, 506.8666 nomin:1, 64.66148 north:1, 10.549808 northern:1, 27.303299 number:1, 5.8547926 older:1, 74.80949 omit:1, 31.469643 opportun:1, 14.800891 organis:1, 24.346794 page:1, 1.5209209 paid:1, 16.3177 pain:1, 57.408154 peac:1, 15.639306 perceiv:1, 111.82617 perspect:1, 110.91033 photograph:1, 66.417885 packet:1, 119.30314 politician:1, 34.269077 poll:1, 38.48965 pollster:1, 1288.183 popul:1, 21.926973 popular:1, 26.004677 portrai:1, 178.0698 pragmat:1, 265.2712 presid:1, 5.118029 privat:1, 10.977386 pro:1, 43.939682 profession:1, 29.52566 promot:1, 16.983023 pundit:1, 962.03094 put:1, 7.017552 race:1, 36.94942 real:1, 11.970332 realtor:1, 847.76086 rebuild:1, 147.11838 regist:1, 20.461607 repres:1, 8.015343 resourc:1, 17.861177 rev:1, 266.20715 richest:1, 424.90347 rumour:1, 134.93996 schmoke:1, 52815.5 seashor:1, 4401.2915 sector:1, 11.133116 see:1, 54.708412 serv:1, 13.570272 shine:1, 223.7945 skill:1, 42.303165 spread:1, 34.59455 straight:1, 71.614235 strong:1, 11.500131 structur:1, 16.850275 style:1, 28.909903 suburb:1, 208.09889 super:1, 107.63297 surg:1, 73.070694 swai:1, 399.81454 swamp:1, 508.81985 talker:1, 4982.594 technolog:1, 15.807817 televis:1, 15.581632 thread:1, 455.30603 tomorrow:1, 47.342686 translat:1, 86.910484 trend:1, 28.887764 turn:1, 8.733588 type:1, 3.209849 understand:1, 17.446981 unemploy:1, 45.48747 union:1, 10.360044 urg:1, 24.178493 vote:1, 14.81958 war:1, 12.918063 week:1, 4.988524 winner:1, 54.359303 womanis:1, 14670.973 work:1, 4.2602763 year:1, 1.985470

, political clubs, labour unions, and churches and many of them will go for Governor Bill Clinton issue' which so absorbs the political pundits in their endless debates over Governor Clinton's by low-skilled job losses, crime and drugs. Mr Clinton has won the support of the black churches Bascom, one of the city's most influential ministers at a recent Clinton rally. Mr Clinton responded, who would then have difficulty winning elsewhere. But Mr Clinton - and the voters - have been distracted


• [FT921-7766](#)

Term Freq Vector:

mr:39, 4.2022786 bush:23, 47.37666 presid:13, 5.118029 buchanan:12, 1205.8334 democrat:9, 14.2563505 hampshir:8, 363.49277 republican:7, 65.123924 economi:6, 11.139456 elect:6, 10.535707 plai:6, 10.0333395 tax:6, 9.731809 battl:5, 29.698324 messag:5, 36.540405 parti:5, 6.676801 campaign:4, 19.610685 candid:4, 27.675604 care:4, 26.284214 novemb:4, 13.215108 pollit:4, 6.0952682 primari:4, 41.028122 reviv:4, 56.407714

And this holds true, by setting the value of k to 0 all the articles that were returned had 0 occurrences of the term "Clinton".

Next we reset the value of k and will be changing the value of b , what we are expecting the length of the returned articles to change, with the highest b value giving us the shortest articles that match our query and the lowest b value to return the longest articles.



Clinton

Search

IR Model: #docs to retrieve:

BM25 parameters: k b

• [FBIS3-39803](#)

Term Freq Vector:

clinton:9, 45.894596 ciller:8, 918.53046 bosnia:5, 54.376095 minist:5, 6.0379205 prime:5, 12.409074 convers:4, 41.817497 herzegovina:4, 110.77077 telephon:4, 19.682306 ankara:3, 226.28749 develop:3, 4.813267 sarajevo:3, 117.9444 statement:3, 11.338665 u.:3, 9.52008 bill:2, 9.10722 discuss:2, 8.72133 feb:2, 12.911747 letter:2, 14.380173 mr:2, 4.2022786 network:2, 12.110869 presid:2, 5.118029 report:2, 2.316825 tansu:2, 1025.5437 televis:2, 15.581632 trt:2, 862.99835 turkish:2, 114.34402 accompani:1, 38.61347 ad:1, 7.020724 admir:1, 85.53117 afternoon:1, 33.87129 ambassador:1, 55.612823 announc:1, 7.383272 approxim:1, 44.83489 articl:1, 6.381227 attack:1, 17.65165 bhutto:1, 762.1284 center:1, 12.6289425 courag:1, 137.07631 decis:1, 7.7518234 departur:1, 71.81874 document:1, 5.88028 express:1, 12.860813 fbi:1, 4.515535 februari:1, 10.864704 feel:1, 14.393497 fulli:1, 26.123009 gmt:1, 6.22494 half:1, 9.219939 hope:1, 8.490556 hour:1, 11.597352 inhuman:1, 818.845 innoc:1, 113.606155 issu:1, 4.2502656 languag:1, 7.2994957 last:1, 74.56657 launch:1, 18.512917 meet:1, 4.9425874 minut:1, 20.530807 nato:1, 70.233376 pakistani:1, 350.4678 palac:1, 116.46196 peopl:1, 3.824605 place:1, 6.585556 presidenti:1, 35.287968 press:1, 9.285752 read:1, 17.73999 regard:1, 23.377966 relai:1, 207.77144 share:1, 6.0378513 state:1, 2.766687 still:1, 1666.1041 stress:1, 20.094164 struggl:1, 26.643545 subject:1, 12.701835 terror:1, 76.6553 text:1, 4.375694 thank:1, 299.57742 turkei:1, 75.69944 type.bfn:1, 9.673877 type.daili:1, 10.209642 visit:1, 12.670449 weu:1, 41.75798

Clinton Telephones Ciller To Discuss Bosnia TA1802194494 Ankara TRT Television Network in Turkish 1800 GMT-read report over stills of Prime Minister Tansu ciller and U.S. President Bill Clinton] [Text] U.S. President Bill Clinton telephoned Prime Minister Tansu Ciller this afternoon and conversed after her conversation with Clinton, Ciller said: Both Mr. Clinton and I are hopeful about developments in Bosnia-Herzegovina following the NATO decision on the subject. The statement added: Mr. Clinton

• [FBIS4-56551](#)

Term Freq Vector:

clinton:9, 45.894596 cuba:4, 130.11949 havana:4, 189.77902 bosnia:3, 54.376095 explan:3, 72.88918 polici:3, 6.303694 africa:2, 26.38664 caribbean:2, 114.567245 countri:2, 2.299005 elect:2, 10.535707 halt:2, 320.28806 intervent:2, 65.79731 note:2, 9.65672 possibl:2, 16.720116 radio:2, 13.676041 south:2, 8.453997 spanish:2, 27.777164 talk:2, 7.774646 u.:2, 9.52008 washington:2, 11.138751 william:2, 23.074621 african:1, 50.290897 agreement:1, 8.577983 announc:1, 7.383272 area:1, 6.401103 atlanta:1, 136.82771 attack:1, 17.65165 base:1, 5.1279674 begin:1, 8.61127 bfn:1, 8.5514555 blockad:1, 194.81926 bosnian:1, 97.15876 capabl:1, 33.372616 case:1, 7.778768 cnn:1, 766.553 contin:1, 122.769646 contribut:1, 14.917526 current:1, 9.070464 daili:1, 6.8684325 decad:1, 24.325489 democraci:1, 37.99403 detail:1, 15.744671 develop:1, 4.813267 differ:1, 18.405819 econom:1, 5.4726553 exclud:1, 37.719967 fact:1, 10.156046 favor:1, 28.54429 fbi:1, 4.515535 forc:1, 6.037989 foreign:1, 5.7644367 forum:1, 72.34009 gave:1, 19.441765 georgia:1, 99.99148 gmt:1, 6.22494 hard:1, 14.552531 head:1, 7.687958 held:1, 8.504089 hous:1, 8.011087 inform:1, 6.424696 iron:1, 59.289963 island:1, 39.30017 israel:1, 49.7696 journalist:1, 50.43497 kei:1, 18.703035 kill:1, 20.667383 lat:1, 43.04091 lift:1, 28.044125 militari:1, 13.252578 morn:1, 17.744164 nation:1, 3.5148706 nato:1, 70.233376 negoti:1, 12.680488 open:1, 6.1597447 organ:1, 9.121691 peac:1, 15.639306 plane:1, 70.00994 plo:1, 138.07974 prais:1, 71.430214 prc:1, 214.26166 presid:1, 5.118029 priorit:1, 37.793583 protect:1, 10.944857 regard:1, 23.377966 reit:1, 95.4727 report:1, 2.316825 restor:1, 36.459686 result:1, 5.1440997 review:1, 10.82773 rwanda:1, 329.27368 salut:1, 464.1081 sanction:1, 51.95308 separ:1, 17.849104 serb:1, 78.99416 serbia:1, 185.38258 sign:1, 8.550071 solut:1, 25.636103 spoke:1, 59.36327 stop:1, 15.191274 support:1, 5.7452493 tension:1, 59.644833 territori:1, 23.961302 text:1, 4.375694 ti:1, 28.150251 topic:1, 73.192215 trade:1, 5.983403 wednesday:1, 11.208485 white:1, 19.304617 work:1, 4.260276

FBIS4-56551 "drlat087_a_94003" FBIS-LAT-94-087 Daily Report 4 May 1994 CARIBBEAN Cuba Clinton Policy 'Explanation' Noted Clinton Policy 'Explanation' Noted PA0405141894 Havana Radio Havana Cuba Information Review] [Text] U.S. President William Clinton gave a detailed explanation of the priorities. In this regard, Clinton talked about current areas of tension such as Bosnia, Rwanda, and Haiti. He also talked about the elections held in South Africa, among other topics. As for Haiti, Clinton

• [FBIS3-11613](#)

Term Freq Vector:

clinton:6, 45.894596 havana:3, 189.77902 cuba:2, 130.11949 dinner:2, 91.53466 mar:2, 11.091033 network:2, 12.110869 radio:2, 13.676041 rebel:2, 582.3098 spanish:2, 27.777164 whitewat:2, 2266.7598 arkansa:1, 326.0216 articl:1, 6.381227 blockad:1, 194.81926 caribbean:1, 114.567245 case:1, 7.778768 cite:1, 33.294773 collect:1, 17.674686 compani:1, 3.3356385 david:1, 14.634791 democrat:1, 14.634791 democrat:1, 5.88028 embargo:1, 129.0386 fact:1, 10.156046 fbi:1, 4.515535 fund:1, 7.965538 garner:1, 463.29385 gmt:1, 6.22494 governor:1, 35.86304 hale:1, 949.91907 hilari:1, 446.45392 judg:1, 25.404282 languag:1, 7.2994957 lat:1, 43.04091 lift:1, 28.044125 million:1, 9.764555 night:1, 12.521456 own:1, 15.593594 parti:1, 6.676801 presid:1, 5.118029 properti:1, 15.284032 prosecut:1, 82.15197 provid:1, 5.7699156 report:1, 2.316825 retain:1, 33.85609 speak:1, 19.411755 special:1, 9.780286 testifi:1, 122.71259 text:1, 4.375694 transact:1, 43.037403 type.bfn:1, 9.673877 type.daili:1, 10.209642 u.:1, 9.52008 wife:1, 38.24162 william:1, 23.074621 worri:1, 26.97283

Cuba Havana Cites Clinton on Retaining Embargo FL2203182294 Havana Radio Rebelde Network in Spanish more than \$3 million -- U.S. President William Clinton said he will not lift the blockade on Cuba. During the dinner, Clinton seemed worried about the fact that he will have to testify Clinton provided some \$100,000 to the Whitewater company, through which Clinton and his wife Hillary owned some properties when Clinton was governor of Arkansas.

• [FBIS3-59184](#)

Term Freq Vector:

clinton:13, 45.894596 presid:8, 5.118029 plan:7, 4.645122 state:7, 2.766687 popular:6, 26.004677 republican:5, 65.123924 american:4, 9.836388 unit:4, 2.9128659 foreign:3, 5.7644367 reform:3, 15.490687 speech:3, 34.701378 union:3, 10.360044 address:2, 14.078502 bill:2, 9.10722 brussel:2, 48.81736 compromis:2, 53.756233 crime:2, 34.630844 de:2, 16.868572 domest:2, 14.97038 dutch:2, 78.81734 fight:2, 18.80492 jan:2, 11.734431 law:2, 8.890151 morgen:2, 7766.9854 opposit:2, 18.34636 privat:2, 10.977386 support:2, 5.7452493 talk:2, 7.774646 theme:2, 75.67775 wing:2, 49.452717 absolut:1, 47.35542 affair:1, 8.156456 anticrim:1, 7882.9106 aris:1, 62.128574 articl:1, 6.381227 averag:1, 13.976052 belgium:1, 85.51733 benefit:1, 12.841426 benelux:1, 546.7443 billion:1, 23.914648 boost:1, 29.499273 brilliantli:1, 772.15643 build:1, 8.077741 bush:1, 47.37666 call:1, 5.006873 camp:1, 53.001003 care:1, 19.872635 chapter:1, 54.398495 coast:1, 38.30819 commentari:1, 109.01032 comprehens:1, 46.884598 consider:1, 15.727793 convers:1, 41.817497 cost:1, 6.4599004 daili:1, 6.8684325 danger:1, 24.248428 death:1, 24.477684 demagog:1, 3000.8806 democrat:1, 14.2563505 document:1, 5.88028 doubt:1, 23.915731 east:1, 9.873901 expand:1, 20.779596 face:1, 9.155687 fanat:1, 631.0096 fbi:1, 4.515535 featur:1, 29.954344 financ:1, 8.422719 find:1, 8.781362 floor:1, 41.066402 found:1, 9.5647335 georg:1, 26.284214 good:1, 5.906387 ground:1, 17.080786 health:1, 15.862416 hobbyhors:1, 40627.31 hunt:1, 88.2612 imag:1, 40.36031 inclin:1, 169.82475 includ:1, 3.5955327 irresist:1, 666.86237 irrevoc:1, 747.0368 isol:1, 67.4011 issu:1, 4.2502656 languag:1, 7.2994957 larg:1, 7.562357 leadership:1, 26.693369 left:1, 9.204353 make:1, 3.423398 manag:1, 5.0954638 million:1, 9.764555 monei:1, 9.500558 natur:1, 12.473549 oblig:1, 34.074516 order:1, 6.43401 origin:1, 16.331324 overtak:1, 782.45184 pass:1, 14.946232 peopl:1, 3.824605 plai:1, 10.0333395 policemen:1, 231.24124 polici:1, 6.303694 poverti:1, 110.8405 predecessor:1, 124.85934 priorit:1, 31.793583 prison:1, 46.484333 provid:1, 5.7699156 question:1, 8.762713 quicker:1, 433.6248 rais:1, 9.425112 reach:1, 9.473123 real:1, 11.970332 recent:1, 5.781474 remain:1, 6.573589 report:1, 2.316825 review:1, 10.82773 rival:1, 39.966328 scarc:1, 160.09549 sector:1, 11.133116 secur:1, 6.0898566 senat:1, 49.99574 sentenc:1, 47.491684 short:1, 12.639521 shown:1, 32.956135 shut:1, 89.320984 significantli:1, 49.217686 skilfulli:1, 2824.3584 social:1, 11.985272 splendid:1, 358.31412 stake:1, 26.012362 stricken:1, 418.50635 strike:1, 27.94471 strikingli:1, 1408.4133 subject:1, 12.701835 system:1, 6.752001 text:1, 4.375694 time:1, 4.46516 topic:1, 73.192215 tradit:1, 22.989248 train:1, 16.756187 turn:1, 8.733588 type.bfn:1, 9.673877 type.daili:1, 10.209642 u.:1, 9.52008 unattribut:1, 137.86348 unclear:1,

By setting the value of b to 1, we received a similar result to just using TF-IDF, the top results are the shortest articles with the most occurrence of the term "Clinton", we noticed that much

affairs" president. With his State of the Union address, Clinton has left

FT921-5527 AN-CCDA0AACFT 920304 FT 04 MAR 92 / Clinton in Georgia Bill Clinton, Democratic

FT921-11595 AN-CBCBKAABFT 920131 FT 31 JAN 92 / World News In Brief: Clinton recovers lead Arkansas governor Bill Clinton regained a

Next we set the value of b to 0.

BM25 parameters: k b

term("Clinton" mentioned 52 times).



This time around we will be searching up with multiple words so it will be different to just searching up "Clinton".

IR Model: #docs to retrieve: BM25 parameters: k b

- [FT941-5673](#)

Term Freq Vector:

world:3, 5.1912227 bomb:2, 57.20297 center:2, 12.6289425 trade:2, 5.983403 _an:1, 2.512583 america:1, 11.858525 began:1, 15.665748 convict:1, 57.684032 countri:1, 2.299005 court:1, 13.315392 ecfayaaact:1, 528155.0 financi:1, 2.1300352 ft:1, 2.4916968 fundamentalist:1, 310.49677 guilt:1, 91.78919 industri:1, 2.6442788 london:1, 2.4773097 mar:1, 11.091033 moslem:1, 150.04404 new:1, 3.8393402 page:1, 1.5209209 peop:1, 27.695595 peopl:1, 3.824605 present:1, 7.5913787 prosecutor:1, 82.15197 septemb:1, 11.705303 state:1, 2.766687 time:1, 1.46516 trial:1, 41.375244 type:1, 3.209849 unit:1, 2.9128659 usz:1, 27.915169 wit:1, 46.607395 york:1, 14.04929

Trade Center Four Moslem fundamentalists were convicted of bombing New York's World Trade CenterFT941-5673 _AN-ECFAYAACT 940305 FT 05 MAR 94 / World News in Brief: Four guilty of bombing World

- [FT942-7648](#)

Term Freq Vector:

bomb:2, 57.20297 center:2, 12.6289425 trade:2, 5.983403 world:2, 5.1912227 year:2, 1.9854704 _an:1, 2.512583 abouhalimi:1, 528155.0 america:1, 11.858525 ayyad:1, 31067.941 convict:1, 57.684032 countri:1, 2.299005 court:1, 13.315392 eeydraact:1, 528155.0 financi:1, 2.1300352 ft:1, 2.4916968 fundamentalist:1, 310.49677 industri:1, 2.6442788 jail:1, 71.848045 london:1, 2.4773097 mahmoud:1, 4475.8896 mohammad:1, 881.7275 moslem:1, 150.04404 new:1, 3.8393402 nidal:1, 3693.3916 page:1, 1.5209209 peop:1, 27.695595 peopl:1, 3.824605 salamah:1, 5030.048 sentenc:1, 47.491684 state:1, 2.766687 time:1, 1.46516 type:1, 3.209849 unit:1, 2.9128659 usz:1, 27.915169 york:1, 14.04929

of bombing the World Trade Center in New York, were each sentenced to 240 years' jail. Countries:- USZFT942-7648 _AN-EEYDRAACT 940525 FT 25 MAY 94 / World News in Brief: 240 years for Trade Center bombing Mohammad Salamah, Nidal Ayyad and Mahmoud Abouhalimi, Moslem fundamentalists convicted

- [FT941-5645](#)

Term Freq Vector:

york:4, 14.049291 bomb:2, 57.20297 guilt:2, 91.78919 moslem:2, 150.04404 peopl:2, 3.824605 reuter:2, 42.908035 _an:1, 2.512583 abouhalima:1, 176051.67 ahmad:1, 285.4892 ajaj:1, 132038.75 america:1, 11.858525 ayyad:1, 31067.941 center:1, 12.6289425 charg:1, 9.539682 conspiraci:1, 182.56308 countri:1, 2.299005 court:1, 13.315392 di:1, 33.435997 face:1, 9.155687 financi:1, 2.1300352 found:1, 9.5647335 ft:1, 2.4916968 fundamentalist:1, 310.49677 industri:1, 2.6442788 juri:1, 116.43629 london:1, 2.4773097 mahmud:1, 1002.19165 mar:1, 11.091033 mohammad:1, 881.7279 nidal:1, 3693.3916 page:1, 1.5209209 peop:1, 27.695595 report:1, 2.316825 salameh:1, 21126.2 state:1, 2.766687 time:1, 1.46516 trade:1, 5.983403 type:1, 3.209849 unit:1, 2.9128659 usz:1, 27.915169 world:1, 5.1912227 year:1, 1.9854704 yesterdai:1, 5.520418

NEW YORK A New York jury yesterday found four Moslem fundamentalists guilty of bombing the World TradeFT941-5645 _AN-ECFAYAA4FT 940305 FT 05 MAR 94 / Moslems guilty of New York bombing By REUTER Center last year, in which six people died, Reuter reports from New York. Mohammad Salameh, Nidal Ayyad

- [FBIS3-28461](#)

Term Freq Vector:

report:6, 2.316825 surin:4, 7235.0 thai:4, 176.99564 thailand:4, 106.375626 confirm:3, 20.810709 foreign:3, 5.7644367 agenc:2, 8.504364 bangkok:2, 163.71823 bomb:2, 57.20297 center:2, 12.6289425 channel:2, 35.836273 check:2, 32.65659 color:2, 48.49463 feb:2, 12.911747 inform:2, 6.424696 make:2, 3.423398 record:2, 8.502173 sourc:2, 12.523534 suspect:2, 42.675743 televis:2, 15.581632 trade:2, 5.983403 world:2, 5.1912227 articl:1, 6.381227 author:1, 6.9454784 bad:1, 23.531076 begin:1, 8.61127 clarif:1, 176.81787 council:1, 10.228822 deputi:1, 12.847048 document:1, 5.88028 ea:1, 28.535036 end:1, 4.1984706 evid:1, 18.504484 face:1, 9.155687 fbi:1, 4.515535 give:1, 8.017168 gmt:1, 6.22494 govern:1, 3.3375988 headlin:1, 95.83651 hide:1, 113.5573 idl:1, 278.41592 languag:1, 7.2994957 matter:1, 12.488296 minist:1, 6.0379205 ministri:1, 14.479521 nation:1, 3.5148706 new:1, 3.8393402 newspap:1, 24.956528 phitsuan:1, 10156.827 progress:1, 18.391079 proof:1, 93.09977 quot:1, 36.404396 read:1, 17.73999 reliabl:1, 78.2336 remain:1, 6.573589 reput:1, 59.21684 secur:1, 6.0898566 speak:1, 19.411755 text:1, 4.375694 true:1, 25.561659 type:bf:1, 9.673877 type:daili:1, 10.209642 u:1, 9.5200

on World Trade Center Bombing Suspects BK0502155094 Bangkok Thai Color Television Channel 9 in Thai the report by a foreign news agency that the suspects in the bombing of the World Trade Center

- [FT933-6564](#)

Term Freq Vector:

court:2, 13.315392 deni:2, 22.6822 egyptian:2, 188.29056 sheikh:2, 1288.183 world:2, 5.1912227 _an:1, 2.512583 abdel:1, 2164.5696 alleg:1, 27.432348 america:1, 11.858525 aug:1, 26.331388 blow:1, 63.457287 bomb:1, 57.20297 center:1, 12.6289425 cleric:1, 422.524 conspir:1, 382.721 conspiraci:1, 182.56308 countri:1, 2.299005 defend:1, 25.924263 financi:1, 2.1300352 ft:1, 2.4916968 hosni:1, 1692.8044 industri:1, 2.6442788 involv:1, 8.572136 london:1, 2.4773097 mubarak:1, 392.38855 murder:1, 50.710995 new:1, 3.8393402 omar:1, 817.5774 page:1, 1.5209209 peop:1, 27.695595 peopl:1, 3.824605 plot:1, 107.392235 presid:1, 5.118029 radic:1, 53.28978 rahman:1, 1109.5693 state:1, 2.766687 time:1, 1.46516 trade:1, 5.983403 type:1, 3.209849 unit:1, 2.9128659 usz:1, 27.915169 york:1, 14.04929

conspiring to bomb the World Trade Center and involvement in alleged plots to blow up the UN and murderFT933-6564 _AN-DHOCBAAMFT 930827 FT 27 AUG 93 / World News in Brief: Sheikh denies conspiracy

- [FBIS3-6734](#)

Term Freq Vector:

police:26, 16.88421 bomb:17, 57.20297 found:10, 9.5647335 tank:10, 68.18423 truck:10, 54.499535 embassi:8, 83.357796 man:7, 13.972355 mideast:7, 685.9156 rent:7, 73.242966 trade:7, 5.983403 water:7, 16.030443 world:7, 5.1912227 center:6, 12.6289425 shop:6, 31.559904 bangkok:5, 163.71823 believ:5, 11.696231 chaiyasit:5, 58683.89 foreign:5, 5.7644367 gen:5, 83.59528 hous:5, 8.011087 investig:5, 17.253765 pol:5, 1415.9651 thailand:5, 106.375626 thursdai:5, 9.649487 bang:4, 311.41214 isra:4, 60.291668 lt:4, 150.85832 metropolitan:4, 125.27396 road:4, 19.250437 year:4, 1.9854704 yesterdai:4, 5.5204186 agenc:3, 8.504364 banyat:3, 24007.045 english:3, 8.705804 fertil:3, 199.52966 involv:3, 8.572136 kapi:3, 88025.836 left:3, 9.204353 lumphini:3, 52815.5 order:3, 6.43401 owner:3, 23.05649 peopl:3, 3.824605 person:3, 7.458447 search:3, 35.899605 secur:3, 6.0898566 shophous:3, 44012.918 station:3, 20.627026 team:3, 14.797159 terrorist:3, 88.52749 assign:2, 48.35256 belong:2, 48.91683 bodi:2, 17.224504 carri:2, 10.239928 central:2, 9.204995 confid:2, 18.553888 countri:2, 2.299005 deliv:2, 28.163761 depart:2, 7.959656 diesel:2, 249.01225 director:2, 6.5301065 driver:2, 65.495415 drove:2, 103.88572 earli:2, 9.261815 earlier:2, 11.460205 explos:2, 63.887142 fuel:2, 34.73333 gener:2, 3.3491335 homemad:2, 851.8629 identifi:2, 20.215685 immigr:2, 95.4727 inform:2, 6.424696 intellig:2, 57.6085: israel:2, 49.7696 item:2, 28.719685 link:2, 16.459066 lot:2, 14.825403 made:2, 3.9773703 makro:2, 25150.238 mar:2, 11.091033 march:2, 7.452974 meet:2, 4.9425874 minist:2, 6.0379205 motorcycl:2, 312.51776 motorcyclist:2, 2357.8347 mr:2, 4.2022786 nation:2, 3.5148706 offic:2, 4.993146 park:2, 19.99224 plot:2, 107.392235 post:2, 13.686317 pratin:2, 33009.688 prime:2, 12.409074 question:2, 8.762713 ramkhamhaeng:2, 88025.836 report:2, 3.316825 respons:2, 7.309497 spoke:2, 59.36327 staff:2, 7.3542805 store:2, 26.132057 superstor:2, 1075.6721 suspect:2, 42.675743 u:2, 9.52008 vehicl:2, 24.81465 wire:2, 16.863726 work:2, 4.2602763 york:2, 14.049291 aboard:1, 167.03194 accid:1, 49.29118 accident:1, 267.14972 account:1, 9.279226 add:1, 22.221264 address:1, 14.078502 admit:1, 27.425226 agent:1, 32.7904 aqo:1, 8.942231 ahmad:1, 285.4892 almal:1,

With a much longer search query, all the returned articles are relevant to what we wanted,

One interesting thing is that the 1st result does not contain the sequence of words “World Trade Center bombing” but has been ranked so high because the words “world”, “bomb”, “center”, and “trade” occurred 3, 2, 2, 2 times. Whereas the first result that does have the complete sequence of words “World Trade Center bombing” ranked 4th with the following words “word”, “bomb”, “center”, and “trade” occurred 2, 2, 2, 2 times.

Switch over the BM-25 with $k=1.2$ and $b=0.75$ we get

	World Trade Center bombing	Search
	IR Model: [BM25] > docs to retrieve: [20]	
	BM25's parameters: k [1.2]	b [0.75]
<p>FTM#5623 Term Freq Vector 105 3 191227 bomb 2 7 20297 center 2 12 6289425 water 2 5 983403 1 21 512683 the 1 11 856526 being 1 15 665748 town 1 57 684032 choice 1 2 299905 count 1 13 315392 egyptair# 1 4293 53 fr# 2 1 130302# 1 2 491668 fundamental# 1 310 49677 unit# 1 19 781919 1 110 427180 border# 1 150 644966 the# 1 3 4393042 place 1 1 5202959 group 1 3 634065 being# 1 7 531377 prosecutor 1 62 151971 soldier# 1 11 795303 state 1 2 766687 time 1 1 44516 1 unit 1 41 372424 city 1 3 2094949 air 1 2 9126655 use 1 27 9151681 unit 1 46 4607395 unit 1 1 140429</p>		
<p>FTM#27648 Term Freq Vector 105 3 191227 bomb 2 7 20297 center 2 12 6289425 water 2 5 983403 1 21 512683 the 1 11 856526 being 1 15 665748 town 1 57 684032 choice 1 2 299905 count 1 13 315392 egyptair# 1 4293 53 fr# 2 1 130302# 1 2 491668 fundamental# 1 310 49677 unit# 1 19 781919 1 110 427180 border# 1 150 644966 the# 1 3 4393042 place 1 1 5202959 group 1 3 634065 being# 1 7 531377 prosecutor 1 62 151971 soldier# 1 11 795303 state 1 2 766687 time 1 1 44516 1 unit 1 41 372424 city 1 3 2094949 air 1 2 9126655 use 1 27 9151681 unit 1 46 4607395 unit 1 1 140429</p>		
<p>FTM#27648 Term Freq Vector 105 3 191227 bomb 2 7 20297 center 2 12 6289425 water 2 5 983403 1 21 512683 the 1 11 856526 being 1 15 665748 town 1 57 684032 choice 1 2 299905 count 1 13 315392 egyptair# 1 4293 53 fr# 2 1 130302# 1 2 491668 fundamental# 1 310 49677 unit# 1 19 781919 1 110 427180 border# 1 150 644966 the# 1 3 4393042 place 1 1 5202959 group 1 3 634065 being# 1 7 531377 prosecutor 1 62 151971 soldier# 1 11 795303 state 1 2 766687 time 1 1 44516 1 unit 1 41 372424 city 1 3 2094949 air 1 2 9126655 use 1 27 9151681 unit 1 46 4607395 unit 1 1 140429</p>		
<p>FTM#27648 Term Freq Vector 105 3 191227 bomb 2 7 20297 center 2 12 6289425 water 2 5 983403 1 21 512683 the 1 11 856526 being 1 15 665748 town 1 57 684032 choice 1 2 299905 count 1 13 315392 egyptair# 1 4293 53 fr# 2 1 130302# 1 2 491668 fundamental# 1 310 49677 unit# 1 19 781919 1 110 427180 border# 1 150 644966 the# 1 3 4393042 place 1 1 5202959 group 1 3 634065 being# 1 7 531377 prosecutor 1 62 151971 soldier# 1 11 795303 state 1 2 766687 time 1 1 44516 1 unit 1 41 372424 city 1 3 2094949 air 1 2 9126655 use 1 27 9151681 unit 1 46 4607395 unit 1 1 140429</p>		
<p>FTM#27648 Term Freq Vector 105 3 191227 bomb 2 7 20297 center 2 12 6289425 water 2 5 983403 1 21 512683 the 1 11 856526 being 1 15 665748 town 1 57 684032 choice 1 2 299905 count 1 13 315392 egyptair# 1 4293 53 fr# 2 1 130302# 1 2 491668 fundamental# 1 310 49677 unit# 1 19 781919 1 110 427180 border# 1 150 644966 the# 1 3 4393042 place 1 1 5202959 group 1 3 634065 being# 1 7 531377 prosecutor 1 62 151971 soldier# 1 11 795303 state 1 2 766687 time 1 1 44516 1 unit 1 41 372424 city 1 3 2094949 air 1 2 9126655 use 1 27 9151681 unit 1 46 4607395 unit 1 1 140429</p>		
<p>FTM#27648 Term Freq Vector 105 3 191227 bomb 2 7 20297 center 2 12 6289425 water 2 5 983403 1 21 512683 the 1 11 856526 being 1 15 665748 town 1 57 684032 choice 1 2 299905 count 1 13 315392 egyptair# 1 4293 53 fr# 2 1 130302# 1 2 491668 fundamental# 1 310 49677 unit# 1 19 781919 1 110 427180 border# 1 150 644966 the# 1 3 4393042 place 1 1 5202959 group 1 3 634065 being# 1 7 531377 prosecutor 1 62 151971 soldier# 1 11 795303 state 1 2 766687 time 1 1 44516 1 unit 1 41 372424 city 1 3 2094949 air 1 2 9126655 use 1 27 9151681 unit 1 46 4607395 unit 1 1 140429</p>		
<p>FTM#27648 Term Freq Vector 105 3 191227 bomb 2 7 20297 center 2 12 6289425 water 2 5 983403 1 21 512683 the 1 11 856526 being 1 15 665748 town 1 57 684032 choice 1 2 299905 count 1 13 315392 egyptair# 1 4293 53 fr# 2 1 130302# 1 2 491668 fundamental# 1 310 49677 unit# 1 19 781919 1 110 427180 border# 1 150 644966 the# 1 3 4393042 place 1 1 5202959 group 1 3 634065 being# 1 7 531377 prosecutor 1 62 151971 soldier# 1 11 795303 state 1 2 766687 time 1 1 44516 1 unit 1 41 372424 city 1 3 2094949 air 1 2 9126655 use 1 27 9151681 unit 1 46 4607395 unit 1 1 140429</p>		
<p>FTM#27648 Term Freq Vector 105 3 191227 bomb 2 7 20297 center 2 12 6289425 water 2 5 983403 1 21 512683 the 1 11 856526 being 1 15 665748 town 1 57 684032 choice 1 2 299905 count 1 13 315392 egyptair# 1 4293 53 fr# 2 1 130302# 1 2 491668 fundamental#</p>		

The 1st and 2nd position has not changed, that might be that the value for b is a bit too high as these two articles are not ideal results, as they are too short

1st:

FT941-5673 _AN-ECFAYAACT 940305 FT 05 MAR 94 / World News in Brief: Four guilty of bombing World Trade Center Four Moslem fundamentalists were convicted of bombing New York's World Trade Center. Prosecutors presented 207 witnesses during the trial, which began in September. Countries:- USZ United States of America. Industries:- P9211 Courts. Types:- PEOP People. The Financial Times London Page 1

2nd:

FT942-7648 _AN-EEYDRAACFT 940525 FT 25 MAY 94 / World News in Brief: 240 years for Trade Center bombing Mohammad Salamah, Nidal Ayyad and Mahmoud Abouhalimi, Moslem fundamentalists convicted of bombing the World Trade Center in New York, were each sentenced to 240 years' jail. Countries:- USZ United States of America. Industries:- P9211 Courts. Types:- PEOP People. The Financial Times London Page 1

We do see a major difference in ranking for the 3rd position as a new article was moved up, This article is a bit longer than all other articles coming up top 10 so the ratio of words played a big role in the way it ranked so high. With further examination, the words “word”, “bomb”, “center”, and “trade” occurred 7, 17, 6, 7 times. So the major contributor is the word “bomb” appearing a staggering 17 times in this article.

And at last we decide to change the values of k and b so more relevant articles are on top and those very short ones are ranked lower, to achieve this we don't need to change the value of k because saturation is not that we are concerned about, so we focused on b , we found that a value of 0.5 is sufficient to filter out the articles that were too short.

World Trade Center bombing

Search

IR Model: BM25 docs to retrieve: 20

BM25 parameters: k: 1.2 b: 0.5

FBIS-6734

Term Freq Vector:

police: 26, 16.88421 bomb: 17, 57.20297 found: 10, 9.564733 tank: 10, 68.18423 truck: 10, 54.499535 embassy: 8, 63.597196 man: 7, 13.972355 mideaast: 7, 685.1956 rent: 7, 73.242966 trade: 7, 5.983403 water: 7, 16.038443 world: 7, 5.912227 center: 6, 12.6289425 shop: 6, 31.559904 bangkok: 5, 163.71283 belief: 5, 11.696231 chajayatsi: 5, 58683.89 forayis: 5, 5.7644367 gens: 5, 63.599286 house: 5, 6.011087 invest: 5, 17.252165 pols: 5, 1415.9651 thailand: 5, 106.375626 thursday: 5, 9.484847 bang: 4, 311.412141 tara: 4, 60.291668 ba: 4, 150.05832 metropolitan: 4, 125.27396 road: 4, 19.250437 year: 4, 1.9854704 yesterday: 4, 5.5204186 apenex: 3, 8.564384 bayuati: 3, 24007.045 english: 3, 3.705804 lehti: 3, 106.62669 invok: 3, 8.572136 kapi: 3, 8.002533 kelle: 3, 9.204353 kunghis: 3, 5.501815 order: 3, 6.433011 owner: 3, 23.065466 peopl: 3, 3.824685 person: 3, 7.450447 search: 3, 39.890605 secu: 3, 6.08898666 shophouse: 3, 4.640129 218 station: 2, 20.627026 team: 3, 14.797159 berretto: 3, 68.527495 assign: 2, 48.35256 belong: 2, 48.91683 bond: 2, 17.244504 cam: 2, 10.235902 center: 2, 2.940995 count: 2, 2.990095 dail: 2, 28.163761 depart: 2, 7.959568 deval: 2, 2.4591025 director: 2, 6.5301065 drive: 2, 103.088572 eclair: 2, 2.261615 earler: 1, 11.460295 explore: 2, 63.881472 heur: 3, 34.733333 genre: 2, 2.3491355 homeland: 2, 0.0513629 identifier: 2, 20.215885 inform: 2, 56.42717 inform: 2, 6.244986 inlag: 2, 57.60853 isreal: 2, 49.75961 men: 2, 28.179565 link: 2, 16.459966 loc: 2, 14.825463 made: 2, 3.3773703 make: 2, 2.915028 mar: 2, 11.091033 march: 2, 7.462974 meet: 2, 4.8425874 mint: 2, 6.0379205 motorcycle: 2, 312.51776 motorcycle: 2, 2357.8347 mnr: 2, 4.202276 nation: 2, 3.5148706 off: 2, 4.993146 park: 2, 19.99224 plot: 2, 13.688317 pratin: 2, 33009.688 prime: 2, 12.449074 question: 2, 8.762713 ramkhanhang: 2, 88025.836 report: 2, 23.16825 report: 2, 7.359487 spoke: 2, 59.36327 staff: 2, 7.3542805 store: 2, 26.132057 station: 2, 1075.6727 support: 2, 42.675743 u: 2, 9.520088 vehicl: 2, 24.81465 wire: 2, 16.863726 work: 2, 4.2682763 york: 2, 14.049291 year: 2, 14.049291 year: 2, 1.674872 account: 1, 8.276226 add: 1, 22.221264 address: 1, 14.078582 admin: 1, 27.425226 agent: 1, 32.7794 agent: 1, 8.842231 alarm: 1, 285.4852 alarm: 1, 66915.375 ammorium: 1, 3052.9192 apen: 1, 84.871446 appear: 1, 12.562886 applier: 1, 16.164537 arab: 1, 32.882888 area: 1, 6.481183 arkr: 1, 19.630395 arkr: 1, 19.487677 arkr: 1, 6.381227 ask: 1, 12.040191 assembl: 1, 27.455164 assist: 1, 12.432955 august: 1, 14.665269 baranthathan: 1, 27.979763 bar: 1, 45.296311 base: 1, 5.1279674 batter: 1, 232.9753 big: 1, 10.837283 blow: 1, 63.452787 bomb: 1, 11.883339 border: 1, 25.780086 bought: 1, 32.6223 brew: 1, 271.95448 built: 1, 9.077741 bunchup: 1, 528155.0 bureat: 1, 46.83886 bur: 1, 147.6110 buyer: 1, 49.20393 call: 1, 7.778769 cent: 1, 27.595747 checkpoint: 1, 604.032839 chilton: 1, 10953711 chosen: 1, 61.408233 chuan: 1, 1146.1631 claim: 1, 11.167484 clear: 1, 10.987092 close: 1, 6.391264 come: 1, 19.888537 commant: 1, 169.810322 commant: 1, 51.886726 comm: 1, 7.398994 conclut: 1, 23.727705 connect: 1, 858.178864 confirm: 1, 20.810709 connect: 1, 22.195182 contact: 1, 15.342194 contr: 1, 10.228822 decid: 1, 10.467427 deput: 1, 12.647488 destrat: 1, 44.60015 detain: 1, 17.789886 discover: 1, 135.18173 destrut: 1, 21.568792 document: 1, 5.888288 draw: 1, 49.28653 eat: 1, 26.556356 electr: 1, 22.635838 establish: 1, 8.158598 explo: 1, 17.882162 extra: 1, 39.195175 farooq: 1, 2654.8918 fu: 1, 4.153535 februari: 1, 10.864704 feder: 1, 7.0993347 flm: 1, 33.317888 flut: 1, 6.781382 flout: 1, 18.425883 gang: 1, 106.61183 gather: 1, 39.323578 govern: 1, 3.3375986 group: 1, 4.1013465 hand: 1, 5.0089328 harpoon: 1, 2433.894 hazamat: 1, 284077.5 head: 1, 7.687768 held: 1, 113.5873 high: 1, 5.9292147 hit: 1, 16.542083 hotel: 1, 23.387888 hour: 1, 11.597352 ident: 55.449345 identif: 1, 119.758888 immedi: 1, 26.390586 incid: 1, 36.959763 increas: 1, 5.30942444 inid: 1, 31.625853 instruct: 1, 42.038252 inter: 1, 1484.6676 kanchanakhi: 1, 31867.941 kansai: 1, 35210.332 kasi: 1, 528165.5 kingdom: 1, 6.5074174 knight: 1, 408.1569 lang: 1, 417.184055 language: 7, 2.994957 larg: 1, 7.562357 larg: 1, 12.873589 level: 1, 6.9997488 alpha: 1, 1714.769 live: 1, 9.7485334 make: 1, 3.4233388 manhat: 1, 3364.0447 mater: 1, 15.561691 men: 1, 20.387382 messag: 1, 38.540465 millio: 1, 9.764456 mri: 1, 921.73645 mriabr: 1, 152.567777 mohammad: 1, 881.7279 moon: 1, 9.900558 mont: 1, 4.334291 mot: 1, 4438.2773 mount: 1, 56.483206 muhammad: 1, 227.26118 name: 1, 19.138133 night: 1, 12.521456 nihat: 1, 1833.8716 offic: 1, 4.0586802 pai: 1, 8.941323 pakistani: 1, 350.4878 percent: 1, 17.624554 phetchaburi: 1, 1320.3875 photograph: 1, 64.176838 phumwicheit: 1, 528155.1 plainchoit: 1, 2179.7632 plant: 1, 16.242067 plastic: 1, 87.39359 pol: 1, 57.171555 problem: 1, 6.402888 provinc: 1, 38.244399 pull: 1, 18882.678 receipt: 1, 68.219588 recep: 1, 16.564371 report: 1, 395.9183 repes: 1, 8.015343 reward: 1, 87.89384 room: 1, 16.811657 safe: 1, 161.219488 sampraphong: 1, 11481.631 seat: 1, 12.248788 senior: 1, 15.686676 show: 1, 6.869087 side: 1, 10.463555 sketch: 1, 393.26358 smel: 1, 302.66753 son: 1, 1853.1754 south: 1, 8.453897 special: 1, 9.780286 steel: 1, 46.959633 steach: 1, 4293.343 step: 1, 12.700865 store: 1, 17.209541 strength: 1, 31.784016 strong: 1, 11.590311 sum: 1, 48627.31 subphahnan: 1, 176057.87 surveil: 1, 181.18524 suspect: 1, 244.17705 syrian: 1, 159.98296 target: 1, 20.238526 text: 1, 4.375694 that: 1, 17.995648 thon: 1, 4936.028 time: 1, 1.465162 tourist: 1, 89.24552 transfer: 1, 23.378287 turn: 1, 8.733588 type: 181, 9.673877 type: 181, 10.209542 unknown: 1, 99.33377 vachan: 1, 95.87653 value: 1, 75.10303 wall: 1, 23.521645 want: 1, 19.752542 watch: 1, 75.692212 wch: 1, 46.607395 wch: 1, 756.6139 youssef: 1, 11961.29

Reportage, Commentary on Bombing Confirmed Link to World Trade Center Bomb

BK190383684 Bangkok BANGKOK to the same gang which planned a bomb in the World Trade Center in New York early last year. Senior belonged to the same group which exploded a bomb at the World Trade Center in New York. The bomb it was too early to say if the bomb had any connection with the World Trade Center bombing. Mr. Banyat Intelligence Agency arrived at the possible World Trade Center link after a two-hour meeting. The high-level

FT941-5667

Term Freq Vector:

tower: 8, 107.364955 york: 6, 14.049291 bomb: 5, 57.20297 explis: 5, 63.887142 off: 5, 4.993146 polic: 5, 16.88421 blast: 4, 122.62712 broker: 4, 37.287498 call: 4, 5.006873 claim: 4, 11.167484 floor: 4, 4.1066402 peopl: 4, 3.824685 smoke: 4, 97.51754 trade: 4, 5.983403 world: 4, 5.912227 build: 3, 8.077741 center: 1, 12.6289425 commo: 3, 22.929365 complex: 3, 27.320246 exchange: 3, 9.574616 financ: 3, 2.1300352 report: 3, 23182516 dealer: 3, 20.627026 team: 3, 14.797159 berretto: 3, 68.527495 assign: 2, 48.35256 belong: 2, 48.91683 bond: 2, 17.244504 cam: 2, 10.235902 center: 2, 2.940995 count: 2, 2.990095 dail: 2, 28.163761 depart: 2, 7.959568 deval: 2, 2.4591025 director: 2, 6.5301065 drive: 2, 103.088572 eclair: 2, 2.261615 earler: 1, 11.460295 explore: 2, 63.881472 heur: 3, 34.733333 genre: 2, 2.3491355 homeland: 2, 0.0513629 identifier: 2, 20.215885 inform: 2, 56.42717 inform: 2, 6.244986 inlag: 2, 57.60853 isreal: 2, 49.75961 men: 2, 28.179565 link: 2, 16.459966 loc: 2, 14.825463 made: 2, 3.3773703 make: 2, 2.915028 mar: 2, 11.091033 march: 2, 7.462974 meet: 2, 4.8425874 mint: 2, 6.0379205 motorcycle: 2, 312.51776 motorcycle: 2, 2357.8347 mnr: 2, 4.202276 nation: 2, 3.5148706 off: 2, 4.993146 park: 2, 19.99224 plot: 2, 13.688317 pratin: 2, 33009.688 prime: 2, 12.449074 question: 2, 8.762713 ramkhanhang: 2, 88025.836 report: 2, 23.16825 report: 2, 7.359487 spoke: 2, 59.36327 staff: 2, 7.3542805 store: 2, 26.132057 station: 2, 1075.6727 support: 2, 42.675743 u: 2, 9.520088 vehicl: 2, 24.81465 wire: 2, 16.863726 work: 2, 4.2682763 york: 2, 14.049291 year: 2, 14.049291 year: 2, 1.674872 account: 1, 8.276226 add: 1, 22.221264 address: 1, 14.078582 admin: 1, 27.425226 agent: 1, 32.7794 agent: 1, 8.842231 alarm: 1, 285.4852 alarm: 1, 66915.375 ammorium: 1, 3052.9192 apen: 1, 84.871446 appear: 1, 12.562886 applier: 1, 16.164537 arab: 1, 32.882888 area: 1, 6.481183 arkr: 1, 19.630395 arkr: 1, 19.487677 arkr: 1, 6.381227 ask: 1, 12.040191 assembl: 1, 27.455164 assist: 1, 12.432955 august: 1, 14.665269 baranthathan: 1, 27.979763 bar: 1, 45.296311 base: 1, 5.1279674 batter: 1, 232.9753 big: 1, 10.837283 blow: 1, 63.452787 bomb: 1, 11.883339 border: 1, 25.780086 bought: 1, 32.6223 brew: 1, 271.95448 built: 1, 9.077741 bunchup: 1, 528155.0 bureat: 1, 46.83886 bur: 1, 147.6110 buyer: 1, 49.20393 call: 1, 7.778769 cent: 1, 27.595747 checkpoint: 1, 604.032839 chilton: 1, 10953711 chosen: 1, 61.408233 chuan: 1, 1146.1631 claim: 1, 11.167484 clear: 1, 10.987092 close: 1, 6.391264 come: 1, 19.888537 commant: 1, 169.810322 commant: 1, 51.886726 comm: 1, 7.398994 conclut: 1, 23.727705 connect: 1, 858.178864 confirm: 1, 20.810709 connect: 1, 22.195182 contact: 1, 15.342194 contr: 1, 10.228822 decid: 1, 10.467427 deput: 1, 12.647488 destrat: 1, 44.60015 detain: 1, 17.789886 discover: 1, 135.18173 destrut: 1, 21.568792 document: 1, 5.888288 draw: 1, 49.28653 eat: 1, 26.556356 electr: 1, 22.635838 establish: 1, 8.158598 explo: 1, 17.882162 extra: 1, 39.195175 farooq: 1, 2654.8918 fu: 1, 4.153535 februari: 1, 10.864704 feder: 1, 7.0993347 flm: 1, 33.317888 flut: 1, 6.781382 flout: 1, 18.425883 gang: 1, 106.61183 gather: 1, 39.323578 govern: 1, 3.3375986 group: 1, 4.1013465 hand: 1, 5.0089328 harpoon: 1, 2433.894 hazamat: 1, 284077.5 head: 1, 7.687768 held: 1, 113.5873 high: 1, 5.9292147 hit: 1, 16.542083 hotel: 1, 23.387888 hour: 1, 11.597352 ident: 55.449345 identif: 1, 119.758888 immedi: 1, 26.390586 incid: 1, 36.959763 increas: 1, 5.30942444 inid: 1, 31.625853 instruct: 1, 42.038252 inter: 1, 1484.6676 kanchanakhi: 1, 31867.941 kansai: 1, 35210.332 kasi: 1, 528165.5 kingdom: 1, 6.5074174 knight: 1, 408.1569 lang: 1, 417.184055 language: 7, 2.994957 larg: 1, 7.562357 larg: 1, 12.873589 level: 1, 6.9997488 alpha: 1, 1714.769 live: 1, 9.7485334 make: 1, 3.4233388 manhat: 1, 3364.0447 mater: 1, 15.561691 men: 1, 20.387382 messag: 1, 38.540465 millio: 1, 9.764456 mri: 1, 921.73645 mriabr: 1, 152.567777 mohammad: 1, 881.7279 moon: 1, 9.900558 mont: 1, 4.334291 mot: 1, 4438.2773 mount: 1, 56.483206 muhammad: 1, 227.26118 name: 1, 19.138133 night: 1, 12.521456 nihat: 1, 1833.8716 offic: 1, 4.0586802 pai: 1, 8.941323 pakistani: 1, 350.4878 percent: 1, 17.624554 phetchaburi: 1, 1320.3875 photograph: 1, 64.176838 phumwicheit: 1, 528155.1 plainchoit: 1, 2179.7632 plant: 1, 16.242067 plastic: 1, 87.39359 pol: 1, 57.171555 problem: 1, 6.402888 provinc: 1, 38.244399 pull: 1, 18882.678 receipt: 1, 68.219588 recep: 1, 16.564371 report: 1, 395.9183 repes: 1, 8.015343 reward: 1, 87.89384 room: 1, 16.811657 safe: 1, 161.219488 sampraphong: 1, 11481.631 seat: 1, 12.248788 senior: 1, 15.686676 show: 1, 6.869087 side: 1, 10.463555 sketch: 1, 393.26358 smel: 1, 302.66753 son: 1, 1853.1754 south: 1, 8.453897 special: 1, 9.780286 steel: 1, 46.959633 steach: 1, 4293.343 step: 1, 12.700865 store: 1, 17.209541 strength: 1, 31.784016 strong: 1, 11.590311 sum: 1, 48627.31 subphahnan: 1, 176057.87 surveil: 1, 181.18524 suspect: 1, 244.17705 syrian: 1, 159.98296 target: 1, 20.238526 text: 1, 4.375694 that: 1, 17.995648 thon: 1, 4936.028 time: 1, 1.465162 tourist: 1, 89.24552 transfer: 1, 23.378287 turn: 1, 8.733588 type: 181, 9.673877 type: 181, 10.209542 unknown: 1, 99.33377 vachan: 1, 95.87653 value: 1, 75.10303 wall: 1, 23.521645 want: 1, 19.752542 watch: 1, 75.692212 wch: 1, 46.607395 wch: 1, 756.6139 youssef: 1, 11961.29

Reportage, Commentary on Bombing Confirmed Link to World Trade Center Bomb

BK190383684 Bangkok BANGKOK to the same gang which planned a bomb in the World Trade Center in New York early last year. Senior belonged to the same group which exploded a bomb at the World Trade Center in New York. The bomb it was too early to say if the bomb had any connection with the World Trade Center bombing. Mr. Banyat Intelligence Agency arrived at the possible World Trade Center link after a two-hour meeting. The high-level

FT941-5673

Term Freq Vector:

tower: 8, 107.364955 york: 6, 14.049291 bomb: 5, 57.20297 explis: 5, 63.887142 off: 5, 4.993146 polic: 5, 16.88421 blast: 4, 122.62712 broker: 4, 37.287498 call: 4, 5.006873 claim: 4, 11.167484 floor: 4, 4.1066402 peopl: 4, 3.824685 smoke: 4, 97.51754 trade: 4, 5.983403 world: 4, 5.912227 build: 3, 8.077741 center: 1, 12.6289425 commo: 3, 22.929365 complex: 3, 27.320246 exchange: 3, 9.574616 financ: 3, 2.1300352 report: 3, 23182516 dealer: 3, 20.627026 team: 3, 14.797159 berretto: 3, 68.527495 assign: 2, 48.35256 belong: 2, 48.91683 bond: 2, 17.244504 cam: 2, 10.235902 center: 2, 2.940995 count: 2, 2.990095 dail: 2, 28.163761 depart: 2, 7.959568 deval: 2, 2.4591025 director: 2, 6.5301065 drive: 2, 103.088572 eclair: 2, 2.261615 earler: 1, 11.460295 explore: 2, 63.881472 heur: 3, 34.733333 genre: 2, 2.3491355 homeland: 2, 0.0513629 identifier: 2, 20.215885 inform: 2, 56.42717 inform: 2, 6.244986 inlag: 2, 57.60853 isreal: 2, 49.75961 men: 2, 28.179565 link: 2, 16.459966 loc: 2, 14.825463 made: 2, 3.3773703 make: 2, 2.915028 mar: 2, 11.091033 march: 2, 7.462974 meet: 2, 4.8425874 mint: 2, 6.0379205 motorcycle: 2, 312.51776 motorcycle: 2, 2357.8347 mnr: 2, 4.202276 nation: 2, 3.5148706 off: 2, 4.993146 park: 2, 19.99224 plot: 2, 13.688317 pratin: 2, 33009.688 prime: 2, 12.449074 question: 2, 8.762713 ramkhanhang: 2, 88025.836 report: 2, 23.16825 report: 2, 7.359487 spoke: 2, 59.36327 staff: 2, 7.3542805 store: 2, 26.132057 station: 2, 1075.6727 support: 2, 42.675743 u: 2, 9.520088 vehicl: 2, 24.81465 wire: 2, 16.863726 work: 2, 4.2682763 york: 2, 14.049291 year: 2, 14.049291 year: 2, 1.674872 account: 1, 8.276226 add: 1, 22.221264 address: 1, 14.078582 admin: 1, 27.425226 agent: 1, 32.7794 agent: 1, 8.842231 alarm: 1, 285.4852 alarm: 1, 66915.375 ammorium: 1, 3052.9192 apen: 1, 84.871446 appear: 1, 12.562886 applier: 1, 16.164537 arab: 1, 32.882888 area: 1, 6.481183 arkr: 1, 19.630395 arkr: 1, 19.487677 arkr: 1, 6.381227 ask: 1, 12.040191 assembl: 1, 27.455164 assist: 1, 12.432955 august: 1, 14.665269 baranthathan: 1, 27.979763 bar: 1, 45.296311 base: 1, 5.1279674 batter: 1, 232.9753 big: 1, 10.837283 blow: 1, 63.452787 bomb: 1, 11.883339 border: 1, 25.780086 bought: 1, 32.6223 brew: 1, 271.95448 built: 1, 9.077741 bunchup: 1, 528155.0 bureat: 1, 46.83886 bur: 1, 147.6110 buyer: 1, 49.20393 call: 1, 7.778769 cent: 1, 27.595747 checkpoint: 1, 604.032839 chilton: 1, 10953711 chosen: 1, 61.408233 chuan: 1, 1146.1631 claim: 1, 11.167484 clear: 1, 10.987092 close: 1, 6.391264 come: 1, 19.888537 commant: 1, 169.810322 commant: 1, 51.886726 comm: 1, 7.398994 conclut: 1, 23.727705 connect: 1, 858.178864 confirm: 1, 20.810709 connect: 1, 22.195182 contact: 1, 15.342194 contr: 1, 10.228822 decid: 1, 10.467427 deput: 1, 12.647488 destrat: 1, 44.60015 detain: 1, 17.789886 discover: 1, 135.18173 destrut: 1, 21.568792 document: 1, 5.888288 draw: 1, 49.28653 eat: 1, 26.556356 electr: 1, 22.635838 establish: 1, 8.158598 explo: 1, 17.882162 extra: 1, 39.195175 farooq: 1, 2654.8918 fu: 1, 4.153535 februari: 1, 10.864704 feder: 1, 7.0993347 flm: 1, 33.317888 flut: 1, 6.781382 flout: 1, 18.425883 gang: 1, 106.61183 gather: 1, 39.323578 govern: 1, 3.3375986 group: 1, 4.1013465 hand: 1, 5.0089328 harpoon: 1, 2433.894 hazamat: 1, 284077.5 head: 1, 7.687768 held: 1, 113.5873 high: 1, 5.9292147 hit: 1, 16.542083 hotel: 1, 23.387888 hour: 1, 11.597352 ident: 55.449345 identif: 1, 119.758888 immedi: 1, 26.390586 incid: 1, 36.959763 increas: 1, 5.30942444 inid: 1, 31.625853 instruct: 1, 42.038252 inter: 1, 1484.6676 kanchanakhi: 1, 31867.941 kansai: 1, 35210.332 kasi: 1, 528165.5 kingdom: 1, 6.5074174 knight: 1, 408.1569 lang: 1, 417.184055 language: 7, 2.994957 larg: 1, 7.562357 larg: 1, 12.873589 level: 1, 6.9997488 alpha: 1, 1714.769 live: 1, 9.7485334 make: 1, 3.4233388 manhat: 1, 3364.0447 mater: 1, 15.561691 men: 1, 20.387382 messag: 1, 38.540465 millio: 1, 9.764456 mri: 1, 921.73645 mriabr: 1, 152.567777 mohammad: 1, 881.7279 moon: 1, 9.900558 mont: 1, 4.334291 mot: 1, 4438.2773 mount: 1, 56.483206 muhammad: 1, 227.26118 name: 1, 19.138133 night: 1, 12.521456 nihat: 1, 1833.8716 offic: 1, 4.0586802 pai: 1, 8.941323 pakistani: 1, 350.4878 percent: 1, 17.624554 phetchaburi: 1, 1320.3875 photograph: 1, 64.176838 phumwicheit: 1, 528155.1 plainchoit: 1, 2179.7632 plant: 1, 16.242067 plastic: 1, 87.39359 pol: 1, 57.171555 problem: 1, 6.402888 provinc: 1, 38.244399 pull: 1, 18882.678 receipt: 1, 68.219588 recep: 1, 16.564371 report: 1, 395.9183 repes: 1, 8.015343 reward: 1, 87.89384 room: 1, 16.811657 safe: 1, 161.219488 sampraphong: 1, 11481.631 seat: 1, 12.248788 senior: 1, 15.686676 show: 1, 6.869087 side: 1, 10.463555 sketch: 1, 393.26358 smel: 1, 302.66753 son: