

File:../data/experiments/base_cased/training_folder/
no_attention/predictions_weak_matching/aquaint
checkpoint=54, opt_thr=0.03, model=local:
pem_similarityglobal: local

Sample: b'APW19981109_0464_htm'

REPEATING Protectively [1 1 Iran 1] will protest to the [1
International Court of Justice 1] at 2 the [2 Hague 2] and
other global bodies about the U.S.-funded [3 3 Radio Free Eur
ope 3] , the Iran Daily reported Monday . It quoted 1 Foreign
Minister [4 4 Kamal Kharrazi 4] as saying the radio `` was
set up to interfere in [2 Iran 2] 's internal affairs . '' It
did not say when the complaints will be filed . The [3
English-language 3] daily also did not say which
organizations [4 Iran 4] will approach besides the [5 5 ICJ 5]
in 6 the [6 Netherlands 6] . On Tuesday , [5 Iran 5] recalled
its ambassador from [7 7 Prague 7] to protest the [8 8 Czech 8
] government 's assistance to the radio and said it would
reduce the level of its economic and political cooperation
with [6 Prague 6] . The American-sponsored station also began
broadcasts directed to [9 9 Iraq 9] on Friday . In a trial
period of several weeks , the station will broadcast one 30-
minute program a day to [7 Iran 7] and [8 Iraq 8] . The [10 10
Farsi 10] - language service to [9 Iran 9] was approved by
the [10 Czech 10] government in August . [11 Radio Free
Europe 11] began transmitting from [11 11 Munich 11] , [12 Ge
rmany 12] , in 1951 , spreading uncensored news to [12 12 Sov
iet 12] - controlled countries behind the [13 13 Iron Curtain
13] during the [14 14 Cold War 14] between the West and the [1
3 2 Communist East 13] . It moved its headquarters to [14
Prague 14] in 1995 following the collapse of communism six
years earlier . ti/vj

1: 14653 Iran

2: 30269 The_Hague

3: 184408 Radio_Free_Europe/Radio_Liberty

4: 186801 Kamal_Kharrazi

5: 14918 International_Court_of_Justice

6: 21148 Netherlands

7: 23844 Prague

8: 5321 Czech_Republic

9: 7515928 Iraq

10: 11600 Persian_language

11: 19058 Munich

12: 26779 Soviet_Union

13: 18953051 Iron_Curtain

14: 325329 Cold_War

1: 317153 Foreign_minister

[winner: 18133296 Ministry_of_Foreign_Affairs_(Israel),
pem_pos=4, final_scores=-0.3952825,
final_local_scores=-0.42350504, pem=-2.1972246,
similarity=3.400327]

gt: 317153 Foreign_minister, pem_pos=14,
final_scores=-0.7893067, final_local_scores=-0.88226396,
pem=-3.5013833, similarity=-3.197737]
[winner: 30876704 Minister_(Catholic_Church), pem_pos=9,
final_scores=-1.6774828, final_local_scores=-1.6836064,
pem=-4.768517, similarity=-20.220522]
gt: 317153 Foreign_minister, pem_pos=6,
final_scores=-2.7355363, final_local_scores=-3.008224,
pem=-4.3779225, similarity=-61.873867]
2: 9209651 Communism
[winner: 97477 Eastern_Bloc, pem_pos=0,
final_scores=0.1824464, final_local_scores=0.16762531,
pem=-0.6931472, similarity=12.861429]
RECALL MISS: 9209651 Communism]
1: 14653 Iran
[pem_pos=0, final_scores=0.6431069,
final_local_scores=0.6937953, pem=-0.13856095,
similarity=25.547466]
2: 30269 The_Hague
[pem_pos=0, final_scores=0.32801276,
final_local_scores=0.32767183, pem=-0.55011845,
similarity=16.859724]
3: 184408 Radio_Free_Europe/Radio_Liberty
[pem_pos=0, final_scores=0.4176342,
final_local_scores=0.43331993, pem=0.0, similarity=17.020844]
4: 186801 Kamal_Kharazi
[pem_pos=0, final_scores=0.3891821,
final_local_scores=0.49499607, pem=0.0, similarity=18.861383]
5: 14918 International_Court_of_Justice
[pem_pos=0, final_scores=0.17468922,
final_local_scores=0.19398871, pem=0.0, similarity=9.878723]
6: 21148 Netherlands
[pem_pos=0, final_scores=0.4529788,
final_local_scores=0.4546821, pem=-0.24940795,
similarity=19.014654]
7: 23844 Prague
[pem_pos=0, final_scores=0.4396726,
final_local_scores=0.44877023, pem=-0.05392049,
similarity=17.775139]
8: 5321 Czech_Republic
[pem_pos=1, final_scores=0.48676878,
final_local_scores=0.48122203, pem=-1.208016,
similarity=25.019716]
9: 7515928 Iraq
[pem_pos=0, final_scores=0.53908795,
final_local_scores=0.53639036, pem=-0.25379938,
similarity=21.476871]
10: 11600 Persian_language
[pem_pos=0, final_scores=0.054758593,
final_local_scores=0.045873687, pem=-0.17037486,
similarity=6.3852]
11: 19058 Munich

[pem_pos=0, final_scores=0.61221725,
final_local_scores=0.6357406, pem=-0.119245164,
similarity=23.709953]
12: 26779 Soviet_Union
[pem_pos=0, final_scores=0.3104555,
final_local_scores=0.2801422, pem=-0.097171314,
similarity=12.978147]
13: 18953051 Iron_Curtain
[pem_pos=0, final_scores=0.21661179,
final_local_scores=0.20137325, pem=0.0, similarity=10.0990925]
14: 325329 Cold_War
[pem_pos=0, final_scores=0.24773954,
final_local_scores=0.24809304, pem=0.0, similarity=11.493304]
1: 14918 International_Court_of_Justice
[pem_pos=0, final_scores=0.25714976,
final_local_scores=0.2840643, pem=0.0, similarity=12.566758]
2: 14653 Iran
[pem_pos=0, final_scores=0.6333119,
final_local_scores=0.6830959, pem=-0.13856095,
similarity=25.228172]
3: 8569916 English_language
[pem_pos=0, final_scores=0.22633056,
final_local_scores=0.24042606, pem=0.0, similarity=11.264506]
4: 14653 Iran
[pem_pos=0, final_scores=0.64126986,
final_local_scores=0.6917887, pem=-0.13856095,
similarity=25.487583]
5: 14653 Iran
[pem_pos=0, final_scores=0.54528207,
final_local_scores=0.5869366, pem=-0.13856095,
similarity=22.358587]
6: 23844 Prague
[pem_pos=0, final_scores=0.07819243,
final_local_scores=0.05390808, pem=-0.05392049,
similarity=5.9916654]
7: 14653 Iran
[pem_pos=0, final_scores=0.5358215,
final_local_scores=0.5766024, pem=-0.13856095,
similarity=22.050196]
8: 7515928 Iraq
[pem_pos=0, final_scores=0.56204045,
final_local_scores=0.56146246, pem=-0.25379938,
similarity=22.225073]
9: 14653 Iran
[pem_pos=0, final_scores=0.5808199,
final_local_scores=0.62575626, pem=-0.13856095,
similarity=23.517044]
10: 5321 Czech_Republic
[pem_pos=1, final_scores=0.5158093,
final_local_scores=0.51294434, pem=-1.208016,
similarity=25.966373]
11: 184408 Radio_Free_Europe/Radio_Liberty

[pem_pos=0, final_scores=0.45737034,
final_local_scores=0.47672564, pem=0.0, similarity=18.316156]
12: 11867 Germany

[pem_pos=0, final_scores=0.81675977,
final_local_scores=0.82926995, pem=-0.15529114,
similarity=29.68128]

13: 97477 Eastern_Bloc

[pem_pos=0, final_scores=0.1824464,
final_local_scores=0.16762531, pem=-0.6931472,
similarity=12.861429]

14: 23844 Prague

[pem_pos=0, final_scores=0.36974669,
final_local_scores=0.3723868, pem=-0.05392049,
similarity=15.495706]

Sample: b'APW20000312_0050_htm'

AP Nicaragua Reporter Garcia Dies [1 MANAGUA 1] , [1 1 Nicar
agua 1] -LRB- AP -RRB- -- Rodolfo Garcia , an [2 Associated
Press 2] reporter who helped chronicle [3 Nicaragua 3] 's
emergence out of 1 civil war , a papal visit , volcanoes and
floods , died in [2 2 Managua 2] after a long illness . He
was 58 . Garcia joined [4 The Associated Press 4] office in [5
Managua 5] in 1986 after working for Radio Nicaragua , where
he was director of 2 short-wave broadcasts , and after
helping found a local news agency , Agencia de Noticias Nueva
Nicaragua . For the [6 AP 6] , he covered the war between the
[3 3 Sandinista 3] government and [4 4 Contra 4] rebels , the
[7 Sandinistas 7] ' loss of power at the ballot box and the
country 's sometimes turbulent effort to put years of war
behind it . He also reported on 3 volcanic eruptions , the
1996 visit of [5 5 Pope John Paul 5] II and the ravages of 6 H
urricane [6 Mitch 6] in 1998 . '' He worked under very
difficult conditions created by a civil war but always
managed to maintain total objectivity and a fair vision under
pressure in reporting for [8 The Associated Press 8] , ''
said Eloy Aguilar , the [9 AP 9] bureau chief for [7 7 Mexico
7] and [8 8 Central America 8] . Garcia was born in
Nandasmo , a town about 15 miles south of [10 Managua 10] .
He attended the [9 9 National Autonomous University of Nicara
gua 9] . Garcia suffered a series of illnesses last year ,
including 4 diabetes and 5 cancer . He died Saturday night .
He was buried in Nandasmo on Sunday . He is survived by his
wife , Ana Leonor Hernandez , and four daughters .

1: 21362 Nicaragua

2: 57042 Managua

3: 29316 Sandinista_National_Liberation_Front

4: 7548 Contras

5: 23805 Pope_John_Paul_II

6: 281923 Hurricane_Mitch

7: 3966054 Mexico

8: 6121 Central_America

9: 6578810 National_Autonomous_University_of_Nicaragua

1: 7085 Civil_war
[winner: 33158 War, pem_pos=0, final_scores=-1.01133,
final_local_scores=-1.0944452, pem=-0.64124906,
similarity=-25.083504]
RECALL MISS: 7085 Civil_war]
[winner: 2941242 Civil_law_(common_law), pem_pos=0,
final_scores=-1.319181, final_local_scores=-1.359942,
pem=-1.0383744, similarity=-30.846823]
gt: 7085 Civil_war, pem_pos=11, final_scores=-1.9595484,
final_local_scores=-2.131582, pem=-3.730266,
similarity=-39.235176]
2: 57980 Shortwave_radio
[winner: 57980 Shortwave_radio, pem_pos=0,
final_scores=-1.4017086, final_local_scores=-1.4816351,
pem=0.0, similarity=-40.125237]
gt: 57980 Shortwave_radio, pem_pos=0,
final_scores=-1.4017086, final_local_scores=-1.4816351,
pem=0.0, similarity=-40.125237]
3: 8921015 Types_of_volcanic_eruptions
[winner: 2380248 Hawaiian_eruption, pem_pos=2,
final_scores=-1.0827746, final_local_scores=-1.0277177,
pem=-1.0802952, similarity=-20.70462]
gt: 8921015 Types_of_volcanic_eruptions, pem_pos=0,
final_scores=-1.1868006, final_local_scores=-1.2153368,
pem=-0.020753756, similarity=-32.065495]
[winner: 1530817 Andisols, pem_pos=4,
final_scores=-1.1672586, final_local_scores=-1.1284251,
pem=-1.385932, similarity=-22.047827]
gt: 8921015 Types_of_volcanic_eruptions, pem_pos=12,
final_scores=-2.3992784, final_local_scores=-2.5397847,
pem=-6.9108415, similarity=-34.1203]
4: 40017873 Diabetes_mellitus
[winner: 515064 Diabetic_neuropathy, pem_pos=2,
final_scores=-1.2508976, final_local_scores=-1.186033,
pem=-1.0986123, similarity=-25.329453]
gt: 40017873 Diabetes_mellitus, pem_pos=0,
final_scores=-1.2580158, final_local_scores=-1.2949405,
pem=0.0, similarity=-34.553894]
5: 105219 Cancer
[winner: 105219 Cancer, pem_pos=0,
final_scores=-1.2570463, final_local_scores=-1.3091271,
pem=-0.023738135, similarity=-34.848156]
gt: 105219 Cancer, pem_pos=0, final_scores=-1.2570463,
final_local_scores=-1.3091271, pem=-0.023738135,
similarity=-34.848156]
1: 21362 Nicaragua
[pem_pos=0, final_scores=0.51284105,
final_local_scores=0.48600626, pem=-0.24729352,
similarity=19.93793]
2: 57042 Managua
[pem_pos=0, final_scores=0.35218585,
final_local_scores=0.3079738, pem=0.0, similarity=13.280266]

3: 29316 Sandinista_National_Liberation_Front
[pem_pos=0, final_scores=0.22984974,
final_local_scores=0.20588318, pem=0.0, similarity=10.233678]
4: 7548 Contras
[pem_pos=2, final_scores=0.04361932,
final_local_scores=0.035887912, pem=-1.3091661,
similarity=12.2801285]
5: 23805 Pope_John_Paul_II
[pem_pos=0, final_scores=0.051342443,
final_local_scores=0.070131704, pem=0.0, similarity=6.182583]
6: 281923 Hurricane_Mitch
[pem_pos=0, final_scores=0.08893578,
final_local_scores=0.09949331, pem=-0.9963309,
similarity=12.476994]
7: 3966054 Mexico
[pem_pos=0, final_scores=0.62084246,
final_local_scores=0.6376465, pem=-0.18258883,
similarity=24.111303]
8: 6121 Central_America
[pem_pos=0, final_scores=0.27798724,
final_local_scores=0.28410363, pem=-0.058720708,
similarity=12.887264]
9: 6578810 National_Autonomous_University_of_Nicaragua
[pem_pos=0, final_scores=0.23539294,
final_local_scores=0.31448418, pem=0.0, similarity=13.474548]
1: 57042 Managua
[pem_pos=0, final_scores=0.32347602,
final_local_scores=0.2766127, pem=0.0, similarity=12.344388]
2: 18935732 Associated_Press
[pem_pos=0, final_scores=0.30277568,
final_local_scores=0.30696586, pem=0.0, similarity=13.250186]
3: 21362 Nicaragua
[pem_pos=0, final_scores=0.52294046,
final_local_scores=0.49703836, pem=-0.24729352,
similarity=20.267149]
4: 18935732 Associated_Press
[pem_pos=0, final_scores=0.27218407,
final_local_scores=0.27354914, pem=0.0, similarity=12.252965]
5: 57042 Managua
[pem_pos=0, final_scores=0.36694735,
final_local_scores=0.3240985, pem=0.0, similarity=13.761458]
6: 18935732 Associated_Press
[pem_pos=1, final_scores=0.15655543,
final_local_scores=0.14724243, pem=-0.68645424,
similarity=12.216766]
7: 29316 Sandinista_National_Liberation_Front
[pem_pos=0, final_scores=0.2911163,
final_local_scores=0.2728075, pem=0.0, similarity=12.230833]
8: 18935732 Associated_Press
[pem_pos=0, final_scores=0.36280745,
final_local_scores=0.37254137, pem=0.0, similarity=15.20709]
9: 18935732 Associated_Press

```
[pem_pos=1, final_scores=0.16340543,
final_local_scores=0.15472502, pem=-0.68645424,
similarity=12.440062]
10: 57042 Managua
[pem_pos=0, final_scores=0.45269054,
final_local_scores=0.4177599, pem=0.0, similarity=16.556501]
```

Sample: b'APW19980627_0596_htm'

Nikhil Chakravarty , a top [1 Indian 1] political columnist , dead at 74 Nikhil Chakravarty , one of [1 1 India 1] 's best known 1 journalist s , has died . He was 74 . Chakravarty was suffering from 2 brain cancer and died in hospital on Saturday , his family said . 3 Politics was his forte and his columns appeared in leading [2 Indian 2] newspapers and magazines . Chakravarty also started a weekly news magazine , `` Mainstream , ' in 1962 and edited it for more than three decades . Last year , the government appointed him chief of [2 2 Prasar Bharati 2] , an autonomous 4 corporation set up to end government control of radio and television . Chakravarty declined a top government award in 1990 , saying no journalist should be identified with the governing establishment . He graduated from [3 3 Calcutta University 3] and studied at [3 4 Oxford University 3] 's [4 Merton College 4] in [4 England 4] . Chakravarty taught history at [5 Calcutta University 5] in the 1930s before taking to 5 journalism as correspondent of a weekly brought out by the [5 5 Communist Party of India 5] . He later joined the CPI and remained a member until 1978 . He closed the Mainstream weekly for a while when Prime Minister [6 6 Indira Gandhi 6] imposed emergency rule in [6 India 6] 1975-77 , jailing political opponents and imposing censorship on newspapers and magazines . Chakravarty was born November 3 , 1913 , in [7 India 7] 's northeastern state of [8 Assam 8] .

1: 14533 India
2: 508128 Prasar_Bharati
3: 519610 University_of_Calcutta
4: 128351 Merton_College,_Oxford
5: 408215 Communist_Party_of_India
6: 15179 Indira_Gandhi

1: 50100 Journalist

```
[winner: 50100 Journalist, pem_pos=0,
final_scores=-1.395542, final_local_scores=-1.5170941,
pem=-0.11764229, similarity=-40.543644]
gt: 50100 Journalist, pem_pos=0, final_scores=-1.395542,
final_local_scores=-1.5170941, pem=-0.11764229,
similarity=-40.543644 ]
```

2: 37284 Brain_tumor

```
[winner: 37284 Brain_tumor, pem_pos=0,
final_scores=-1.2280394, final_local_scores=-1.2573795,
pem=0.0, similarity=-33.433002]
gt: 37284 Brain_tumor, pem_pos=0,
final_scores=-1.2280394, final_local_scores=-1.2573795,
```

```

pem=0.0, similarity=-33.433002 ]
3: 22986 Politics
    [winner: 22986 Politics, pem_pos=0,
    final_scores=-0.34446323, final_local_scores=-0.3904897,
    pem=-1.8374861, similarity=2.4292572]
    gt: 22986 Politics, pem_pos=0, final_scores=-0.34446323,
    final_local_scores=-0.3904897, pem=-1.8374861,
    similarity=2.4292572 ]
4: 7485 Corporation
    [winner: 7485 Corporation, pem_pos=0,
    final_scores=-1.3308773, final_local_scores=-1.4332876,
    pem=-0.04947952, similarity=-38.413372]
    gt: 7485 Corporation, pem_pos=0,
    final_scores=-1.3308773, final_local_scores=-1.4332876,
    pem=-0.04947952, similarity=-38.413372 ]
5: 15928 Journalism
    [winner: 4269660 Carleton_School_of_Journalism,
    pem_pos=1, final_scores=-1.0522875,
    final_local_scores=-1.0332887, pem=-0.6917592,
    similarity=-22.98379]
    gt: 15928 Journalism, pem_pos=0,
    final_scores=-1.5495068, final_local_scores=-1.6998723,
    pem=-0.026584832, similarity=-46.493298 ]
1: 14533 India
    [pem_pos=0, final_scores=0.6518895,
    final_local_scores=0.6251242, pem=-0.13622136,
    similarity=23.48546]
2: 508128 Prasar_Bharati
    [pem_pos=0, final_scores=0.27440524,
    final_local_scores=0.35563514, pem=0.0, similarity=14.702575]
3: 519610 University_of_Calcutta
    [pem_pos=0, final_scores=0.23349528,
    final_local_scores=0.20898351, pem=0.0, similarity=10.326198]
4: 128351 Merton_College,_Oxford
    [pem_pos=0, final_scores=0.07907696,
    final_local_scores=0.11508614, pem=0.0, similarity=7.524113]
5: 408215 Communist_Party_of_India
    [pem_pos=0, final_scores=0.29965597,
    final_local_scores=0.31423748, pem=0.0, similarity=13.467186]
6: 15179 Indira_Gandhi
    [pem_pos=0, final_scores=0.50984716,
    final_local_scores=0.540661, pem=0.0, similarity=20.224115]
1: 14533 India
    [pem_pos=0, final_scores=0.4595465,
    final_local_scores=0.41501868, pem=-0.618294,
    similarity=19.837078]
2: 14533 India
    [pem_pos=0, final_scores=0.28387278,
    final_local_scores=0.22312182, pem=-0.618294,
    similarity=14.110492]
3: 31797 University_of_Oxford
    [pem_pos=0, final_scores=0.35068917,

```


final_local_scores=0.375857, pem=-0.12160307,
 similarity=15.967333]
 4: 9316 England
 [pem_pos=0, final_scores=0.44326663,
 final_local_scores=0.46691132, pem=-0.19053856,
 similarity=19.059456]
 5: 519610 University_of_Calcutta
 [pem_pos=0, final_scores=0.24613367,
 final_local_scores=0.22278902, pem=0.0, similarity=10.738182]
 6: 14533 India
 [pem_pos=0, final_scores=0.6623841,
 final_local_scores=0.6365879, pem=-0.13622136,
 similarity=23.827559]
 7: 14533 India
 [pem_pos=0, final_scores=0.71287715,
 final_local_scores=0.69174397, pem=-0.13622136,
 similarity=25.473526]
 8: 186162 Assam
 [pem_pos=0, final_scores=0.4347509,
 final_local_scores=0.4031213, pem=0.0, similarity=16.119656]

Sample: b'APW19981109_1172_htm'

[1 LONDON 1] : Nov. 26 . Word of the agreement leaked out
 when former captain [1 1 Courtney Walsh 1] , head of the [2 2
 West Indies 2] players association , told the [3 3 Caribbean N
 ews Agency 3] that [4 4 Lara 4] and 1 Hooper had been
 reinstated and the tour was going ahead . The crisis came to
 a head last Wednesday when the [5 5 West Indies Cricket Board
 5] fired superstar 2 batsman [2 Lara 2] as captain and Hooper
 as vice-captain . The two had refused to travel to [3 South
 Africa 3] , demanding better pay and conditions for the
 tour . Team members had insisted that [4 Lara 4] and Hooper
 be reinstated as part of any settlement . [5 Jimmy Adams 5] ,
 3 treasurer of the players ' association , 4 negotiated on
 behalf of the players Monday with Rousseau . [6 6 Ali Bacher 6
] , head of [6 5 South African 6] cricket , also was directly
 involved in Monday 's talks after sitting on the sidelines
 Sunday . [7 West Indies 7] tour manager [7 7 Clive Lloyd 7]
 and [8 WICB 8] member [8 8 Joel Garner 8] were also involved
 in the talks which began Friday . Progress was reported over
 the weekend with the announcement of a 6 sponsorship deal
 offered by former [9 South African 9] 7 wicketkeeper 8 David R
 ichardson , who runs a sports 9 marketing firm in [9 9 Johann
 esburg 9] . Earnings from the sponsorship proposal would
 apparently help meet the players ' demands for better pay ,
 which is reported to range from dlr\$ 15,000 to 60,000 for the
 [10 South African 10] tour . -LRB- jp/sw -RRB-

1: 252677 Courtney_Walsh
 2: 3379941 West_Indies_cricket_team
 3: 6834818 Caribbean_News_Agency
 4: 4367 Brian_Lara
 5: 8381696 West_Indies_Cricket_Board

6: 3034646 Ali_Bacher
7: 624962 Clive_Lloyd
8: 87055 Joel_Garner
9: 58893 Johannesburg
1: 2112133 Carl_Hooper
[winner: 2112133 Carl_Hooper, pem_pos=15,
final_scores=-0.03634657, final_local_scores=-0.07013661,
pem=-3.4250882, similarity=20.622858]
gt: 2112133 Carl_Hooper, pem_pos=15,
final_scores=-0.03634657, final_local_scores=-0.07013661,
pem=-3.4250882, similarity=20.622858]
2: 673526 Batting_(cricket)
[winner: 673526 Batting_(cricket), pem_pos=0,
final_scores=-1.3918016, final_local_scores=-1.5221832,
pem=-0.0001237469, similarity=-41.3346]
gt: 673526 Batting_(cricket), pem_pos=0,
final_scores=-1.3918016, final_local_scores=-1.5221832,
pem=-0.0001237469, similarity=-41.3346]
3: 247097 Treasurer
[winner: 4429258 Treasurer_of_Maryland, pem_pos=2,
final_scores=-1.6123492, final_local_scores=-1.6055858,
pem=-2.434509, similarity=-30.584932]
gt: 247097 Treasurer, pem_pos=0,
final_scores=-1.8641787, final_local_scores=-1.9648516,
pem=-0.5250262, similarity=-51.6902]
4: 22083 Negotiation
[winner: 18401482 Crisis_negotiation, pem_pos=3,
final_scores=-1.2518911, final_local_scores=-1.2328131,
pem=-2.0053334, similarity=-21.794577]
gt: 22083 Negotiation, pem_pos=0,
final_scores=-1.585637, final_local_scores=-1.633537,
pem=-1.6486586, similarity=-35.69263]
5: 11759174 Cricket_South_Africa
[winner: 17416221 South_Africa, pem_pos=0,
final_scores=0.29898864, final_local_scores=0.31871766,
pem=-0.1389451, similarity=14.356489]
RECALL MISS: 11759174 Cricket_South_Africa
[winner: 330091 Rhaphidophoridae, pem_pos=10,
final_scores=-1.5490001, final_local_scores=-1.5371301,
pem=-2.56456, similarity=-27.834839]
RECALL MISS: 11759174 Cricket_South_Africa
6: 545607 Sponsor_(commercial)
[winner: 29858442 Sports_marketing, pem_pos=1,
final_scores=-1.2923204, final_local_scores=-1.309823,
pem=-0.69121295, similarity=-31.2391]
gt: 545607 Sponsor_(commercial), pem_pos=0,
final_scores=-1.4395573, final_local_scores=-1.5018275,
pem=0.0, similarity=-40.727814]
7: 463313 Wicket-keeper
[winner: 463313 Wicket-keeper, pem_pos=0,
final_scores=-1.7561353, final_local_scores=-1.952558,
pem=0.0, similarity=-54.178516]

gt: 463313 Wicket-keeper, pem_pos=0,
final_scores=-1.7561353, final_local_scores=-1.952558,
pem=0.0, similarity=-54.178516]
8: 3182138 Dave_Richardson
[winner: 3182138 Dave_Richardson, pem_pos=11,
final_scores=-0.12818325, final_local_scores=-0.119573295,
pem=-2.8266559, similarity=15.893202]
gt: 3182138 Dave_Richardson, pem_pos=11,
final_scores=-0.12818325, final_local_scores=-0.119573295,
pem=-2.8266559, similarity=15.893202]
9: 59252 Marketing
[winner: 29858442 Sports_marketing, pem_pos=0,
final_scores=-0.95399433, final_local_scores=-0.9402533,
pem=0.0, similarity=-23.969322]
RECALL MISS: 59252 Marketing]
1: 252677 Courtney_Walsh
[pem_pos=0, final_scores=0.76559895,
final_local_scores=0.8019048, pem=0.0, similarity=28.020151]
2: 3379941 West_Indies_cricket_team
[pem_pos=1, final_scores=0.30810553,
final_local_scores=0.25183368, pem=-1.1020415,
similarity=17.598005]
3: 6834818 Caribbean_News_Agency
[pem_pos=0, final_scores=0.18938862,
final_local_scores=0.33620557, pem=0.0, similarity=14.122758]
4: 4367 Brian_Lara
[pem_pos=4, final_scores=0.2529325,
final_local_scores=0.2117593, pem=-2.6667001,
similarity=24.910963]
5: 8381696 West_Indies_Cricket_Board
[pem_pos=0, final_scores=0.30936652,
final_local_scores=0.3022953, pem=0.0, similarity=13.110807]
6: 3034646 Ali_Bacher
[pem_pos=0, final_scores=0.6858032,
final_local_scores=0.73482263, pem=0.0, similarity=26.018286]
7: 624962 Clive_Lloyd
[pem_pos=0, final_scores=0.79382116,
final_local_scores=0.82479197, pem=0.0, similarity=28.70315]
8: 87055 Joel_Garner
[pem_pos=0, final_scores=0.69973564,
final_local_scores=0.7459763, pem=0.0, similarity=26.351135]
9: 58893 Johannesburg
[pem_pos=0, final_scores=0.44490582,
final_local_scores=0.47033036, pem=-0.09562059,
similarity=18.64531]
1: 17867 London
[pem_pos=0, final_scores=0.5534121,
final_local_scores=0.6173535, pem=-0.09086737,
similarity=23.006924]
2: 4367 Brian_Lara
[pem_pos=4, final_scores=0.2795813,
final_local_scores=0.24086908, pem=-2.6667001,

```

similarity=25.779655]
3: 17416221 South_Africa
    [pem_pos=0, final_scores=0.45508057,
final_local_scores=0.48922443, pem=-0.26251382,
similarity=20.116737]
4: 4367 Brian_Lara
    [pem_pos=4, final_scores=0.32834214,
final_local_scores=0.29413292, pem=-2.6667001,
similarity=27.369156]
5: 1026319 Jimmy_Adams
    [pem_pos=0, final_scores=0.41670603,
final_local_scores=0.43372667, pem=0.0, similarity=17.032982]
6: 17416221 South_Africa
    [pem_pos=0, final_scores=0.29898864,
final_local_scores=0.31871766, pem=-0.1389451,
similarity=14.356489]
7: 3379941 West_Indies_cricket_team
    [pem_pos=1, final_scores=0.37468737,
final_local_scores=0.32456422, pem=-1.1020415,
similarity=19.768429]
8: 3379941 West_Indies_cricket_team
    [pem_pos=2, final_scores=0.20218687,
final_local_scores=0.13671067, pem=-2.0165277,
similarity=19.135622]
9: 759263 South_Africa_national_cricket_team
    [pem_pos=1, final_scores=0.07793827,
final_local_scores=0.006619036, pem=-2.617051,
similarity=18.519167]
10: 17416221 South_Africa
    [pem_pos=0, final_scores=0.2864769,
final_local_scores=0.3050505, pem=-0.1389451,
similarity=13.948633]

```

Sample: b'APW19980614_0031_htm'

Spirited bidders snap up art seized from failed finance firms Paintings and other art objects once displayed in elegant offices were snapped up this weekend in [1 1 Thailand 1] at an 1 auction of assets of collapsed 2 financial firms , [1 Thai 1] media reported Sunday . The three-day auction of more than 1,000 items , which ended Sunday , was expected to bring in more than an initial estimate of 30 million 3 baht - LRB- dlrs 697,000 -RRB- . The paintings , sculptures and other art-related objects were seized by [2 Thailand 2] 's Financial Restructuring Authority , a government agency set up to 4 liquidate the assets of 56 finance companies which were shut down last year . The closings were required by the [2 2 International Monetary Fund 2] as part of its dlrs 17.2 billion bailout program for the depressed [3 5 Thai 3] economy . The proceeds of the auction will go to pay the debts of the failed companies . `` The going prices have been very high , much higher than I expected , '' an art gallery owner , Hongjorn Samae-ngamjaroen , was quoted as saying by the [3 3 B

angkok Post 3] . An untitled painting of a lotus pond by the late Thawee Nandakwang was sold for 2.8 million baht -LRB- dlr 65,000 -RRB- , way above its starting bid of 400,000 baht -LRB- dlr 9,300 -RRB- . Prime Minister [4 4 Chuan Leekpai 4] donated two of his sketches to the auction , which was handled by [5 5 Christies 5] , the [4 British 4] auction house . It was held at [5 Bangkok 5] 's 6 Queen Sirikit National Convention Center .

1: 30128 Thailand

2: 15251 International_Monetary_Fund

3: 1307628 Bangkok_Post

4: 1024311 Chuan_Leekpai

5: 155688 Christie's

1: 62912 Auction

[winner: 62912 Auction, pem_pos=0, final_scores=-1.8060923, final_local_scores=-1.9353817, pem=-0.038607303, similarity=-53.455986|
gt: 62912 Auction, pem_pos=0, final_scores=-1.8060923, final_local_scores=-1.9353817, pem=-0.038607303, similarity=-53.455986]

2: 208836 Financial_institution

[winner: 214491 Multinational_corporation, pem_pos=0, final_scores=-1.274273, final_local_scores=-1.3469455, pem=0.0, similarity=-36.105827|

RECALL MISS: 208836 Financial_institution]

[winner: 160463 Central_business_district, pem_pos=2, final_scores=-1.2984419, final_local_scores=-1.3145916, pem=-1.0967067, similarity=-29.176266|

gt: 208836 Financial_institution, pem_pos=12, final_scores=-2.2715278, final_local_scores=-2.4384024, pem=-5.5970454, similarity=-38.23948]

3: 89761 Thai_baht

[winner: 89761 Thai_baht, pem_pos=0, final_scores=-1.2990768, final_local_scores=-1.4345849, pem=0.0, similarity=-38.72116|

gt: 89761 Thai_baht, pem_pos=0, final_scores=-1.2990768, final_local_scores=-1.4345849, pem=0.0, similarity=-38.72116]

4: 565034 Liquidation

[winner: 565034 Liquidation, pem_pos=0, final_scores=-1.4885821, final_local_scores=-1.5647349, pem=-0.029434502, similarity=-42.44503|

gt: 565034 Liquidation, pem_pos=0, final_scores=-1.4885821, final_local_scores=-1.5647349, pem=-0.029434502, similarity=-42.44503]

5: 30133 Economy_of_Thailand

[winner: 30128 Thailand, pem_pos=2, final_scores=0.6057938, final_local_scores=0.56822616, pem=-1.7504287, similarity=30.565817|

RECALL MISS: 30133 Economy_of_Thailand]

[winner: 112685 Economy_of_Indiana, pem_pos=1, final_scores=-1.19803, final_local_scores=-1.1211576, pem=-2.3978953, similarity=-16.327742|

RECALL MISS: 30133 Economy_of_Thailand]
6: 2221039 Queen_Sirikit_National_Convention_Center
[winner: 846986 Sirikit, pem_pos=0,
final_scores=-0.8192023, final_local_scores=-0.90010333,
pem=0.0, similarity=-22.771166]
RECALL MISS: 2221039
Queen_Sirikit_National_Convention_Center]
[winner: 17495368
Convention_Center_(Baltimore_Light_Rail_station), pem_pos=2,
final_scores=-0.8246202, final_local_scores=-0.7115671,
pem=-1.8345186, similarity=-7.168473]
RECALL MISS: 2221039
Queen_Sirikit_National_Convention_Center]
[winner: 42010 Queen_(band), pem_pos=0,
final_scores=-1.5718397, final_local_scores=-1.635771,
pem=-0.43144405, similarity=-42.378696]
RECALL MISS: 2221039
Queen_Sirikit_National_Convention_Center]
1: 30128 Thailand
[pem_pos=0, final_scores=0.66843206,
final_local_scores=0.6366489, pem=-0.24029273,
similarity=24.395334]
2: 15251 International_Monetary_Fund
[pem_pos=0, final_scores=0.40147197,
final_local_scores=0.48512095, pem=0.0, similarity=18.56669]
3: 1307628 Bangkok_Post
[pem_pos=0, final_scores=0.33157176,
final_local_scores=0.37969083, pem=0.0, similarity=15.4204445]
4: 1024311 Chuan_Leekpai
[pem_pos=0, final_scores=0.59105617,
final_local_scores=0.70122075, pem=0.0, similarity=25.01554]
5: 155688 Christie's
[pem_pos=0, final_scores=0.18880539,
final_local_scores=0.27204865, pem=0.0, similarity=12.208187]
1: 30128 Thailand
[pem_pos=2, final_scores=0.5901099,
final_local_scores=0.5510939, pem=-1.7504287,
similarity=30.054554]
2: 30128 Thailand
[pem_pos=0, final_scores=0.7185479,
final_local_scores=0.69139284, pem=-0.24029273,
similarity=26.029003]
3: 30128 Thailand
[pem_pos=2, final_scores=0.6057938,
final_local_scores=0.56822616, pem=-1.7504287,
similarity=30.565817]
4: 31717 United_Kingdom
[pem_pos=0, final_scores=0.54090005,
final_local_scores=0.5592065, pem=-0.48936236,
similarity=23.438778]
5: 56615 Bangkok
[pem_pos=0, final_scores=0.52671164,

final_local_scores=0.51149064, pem=-0.0005963434,
similarity=19.356857]

Sample: b'APW19981120_1056_htm'

Prosecutor : [1 Botha 1] may still stand trial A 1 prosecutor on Friday left open the possibility that former apartheid President [1 1 P.W. Botha 1] would be put on trial for abuses committed during 2 white rule . Prosecutor Jan D'Oliveira said he currently has insufficient evidence to prosecute [2 Botha 2] , but that information collected by the 3 Truth and Reconciliation Commission -LRB- South Africa -RRB- , which investigated apartheid-era abuses , still must be evaluated to ascertain whether charges would ultimately be brought . Last month , the Truth Commission issued a landmark report on its findings . It said [3 Botha 3] had ordered the 1987 bombing of the [2 2 African National Congress 2] ' [4 London 4] headquarters and the 1988 bombing of a [3 3 Johannesburg 3] building housing an anti-apartheid group . When [5 Botha 5] was head of state in the 1980s , thousands of people were detained without trial . Many were tortured and others killed . The commission can grant 4 amnesty to those who fully confess to politically motivated abuses committed during apartheid . [6 Botha 6] has said he has nothing to confess and has not sought amnesty . Earlier this year , [7 Botha 7] was convicted of 5 contempt of court for refusing to testify before the commission . He was handed a suspended one-year jail sentence and a 10,000 6 rand -LRB- dlr 5,700 -RRB- fine . He has appealed the decision . -LRB- sapa-aos -RRB-

1: 383252 P.W. Botha

2: 2503 African_National_Congress

3: 58893 Johannesburg

1: 440936 Prosecutor

[winner: 21217228 New_York_County_District_Attorney,
pem_pos=3, final_scores=-1.3277291,
final_local_scores=-1.3536885, pem=-1.0977756,
similarity=-30.337175]

gt: 440936 Prosecutor, pem_pos=0,
final_scores=-1.3600103, final_local_scores=-1.4319143,
pem=-0.027151007, similarity=-38.493816]

2: 888454 Dominant_minority

[winner: 888454 Dominant_minority, pem_pos=0,
final_scores=-0.79353595, final_local_scores=-0.8391215,
pem=0.0, similarity=-20.951347]

gt: 888454 Dominant_minority, pem_pos=0,
final_scores=-0.79353595, final_local_scores=-0.8391215,
pem=0.0, similarity=-20.951347]

3: 30865 Truth_and_Reconciliation_Commission_(South_Africa)

[winner: 30865
Truth_and_Reconciliation_Commission_(South_Africa),
pem_pos=0, final_scores=-0.63329315,
final_local_scores=-0.70436555, pem=-0.44964123,
similarity=-14.484742]

gt: 30865
Truth_and_Reconciliation_Commission_(South_Africa),
pem_pos=0, final_scores=-0.63329315,
final_local_scores=-0.70436555, pem=-0.44964123,
similarity=-14.484742]
[winner: 17416221 South_Africa, pem_pos=0,
final_scores=-1.0639789, final_local_scores=-1.1351802,
pem=-0.26251382, similarity=-28.358738]
RECALL MISS: 30865
Truth_and_Reconciliation_Commission_(South_Africa)]
4: 18940624 Amnesty
[winner: 18940624 Amnesty, pem_pos=0,
final_scores=-1.3368546, final_local_scores=-1.4385927,
pem=-0.052021187, similarity=-38.55786]
gt: 18940624 Amnesty, pem_pos=0,
final_scores=-1.3368546, final_local_scores=-1.4385927,
pem=-0.052021187, similarity=-38.55786]
5: 7201 Contempt_of_court
[winner: 7201 Contempt_of_court, pem_pos=0,
final_scores=-1.0258298, final_local_scores=-1.0642015,
pem=-0.0049575465, similarity=-27.641222]
gt: 7201 Contempt_of_court, pem_pos=0,
final_scores=-1.0258298, final_local_scores=-1.0642015,
pem=-0.0049575465, similarity=-27.641222]
6: 298686 South_African_rand
[winner: 22971263 Rand, _California, pem_pos=0,
final_scores=-0.87985814, final_local_scores=-0.7619701,
pem=-1.5304596, similarity=-10.326118]
gt: 298686 South_African_rand, pem_pos=2,
final_scores=-1.3984088, final_local_scores=-1.4875264,
pem=-1.6144447, similarity=-31.521439]
1: 383252 P._W._Botha
[pem_pos=0, final_scores=0.76039475,
final_local_scores=0.7065755, pem=0.0, similarity=25.175335]
2: 2503 African_National_Congress
[pem_pos=0, final_scores=0.40676063,
final_local_scores=0.39419782, pem=0.0, similarity=15.853363]
3: 58893 Johannesburg
[pem_pos=0, final_scores=0.4062912,
final_local_scores=0.46249378, pem=-0.09562059,
similarity=18.41145]
1: 383252 P._W._Botha
[pem_pos=0, final_scores=0.5072787,
final_local_scores=0.43073267, pem=0.0, similarity=16.943634]
2: 383252 P._W._Botha
[pem_pos=0, final_scores=0.8428269,
final_local_scores=0.79726815, pem=0.0, similarity=27.881784]
3: 383252 P._W._Botha
[pem_pos=0, final_scores=0.6448911,
final_local_scores=0.5810533, pem=0.0, similarity=21.429502]
4: 17867 London
[pem_pos=0, final_scores=0.26111674,


```

final_local_scores=0.3736636, pem=-0.09086737,
similarity=15.73473]
5: 383252 P._W._Botha
    [pem_pos=0, final_scores=0.6806733,
final_local_scores=0.62013996, pem=0.0, similarity=22.595926]
6: 383252 P._W._Botha
    [pem_pos=0, final_scores=0.6555039,
final_local_scores=0.5926462, pem=0.0, similarity=21.775457]
7: 383252 P._W._Botha
    [pem_pos=0, final_scores=0.5086501,
final_local_scores=0.43223077, pem=0.0, similarity=16.98834]

```

Sample: b'APW19980620_0458_htm'

[1 Iranian 1] POW negotiator holds talks with [2 Iraqi 2] ministers The head of [1 1 Iran 1] 's 1 prisoner of war commission met with two [3 2 Iraqi 3] Cabinet ministers Saturday in a bid to glean information about thousands of [4 Iranian 4] POWs allegedly in [5 Iraq 5] , the official [2 2 Iraqi News Agency 2] reported . [6 Iraqi 6] Foreign Minister [3 3 Mohammed Saeed al-Sahhaf 3] told Abdullah al-Najafi that the two states needed to `` speed up the closure of what remains from the [7 POW 7] and 3 Missing-In-Action file , '' INA said . The issue of [8 POWs 8] and missing persons remains a stumbling block to normalizing relations between the two neighbors . [9 Iraq 9] has long maintained that it has released all [10 Iranian 10] prisoners captured in the 1980-88 [4 4 Iran-Iraq War 4] . The countries accuse each other of hiding [11 POWs 11] and preventing visits by the [5 5 International Committee of the Red Cross 5] to prisoner camps . The [12 ICRC 12] representative in [6 6 Baghdad 6] , Manuel Bessler , told [7 The 7 Associated Press 7] that his organization has had difficulty visiting [13 POWs 13] on both sides on a regular basis . In April , [14 Iran 14] released 5,584 [15 Iraqi 15] POWs and [16 Iraq 16] freed three [17 POWs 17] and 316 [18 Iranians 18] it classified as civil law detainees in the largest exchange since 1990 . More than 1 million people were killed or wounded in the [19 Iran-Iraq War 19] .

```

1: 14653 Iran
2: 8550616 National_Iraqi_News_Agency
3: 202988 Muhammad_Saeed_al-Sahhaf
4: 14889 Iran-Iraq_War
5: 5023832 International_Committee_of_the_Red_Cross
6: 4492 Baghdad
7: 18935732 Associated_Press

```

```

1: 25008 Prisoner_of_war
    [winner: 32611 Vietnam_War, pem_pos=0,
final_scores=-0.9118657, final_local_scores=-1.0147514,
pem=-0.12429769, similarity=-25.516544]
    RECALL MISS: 25008 Prisoner_of_war]
    [winner: 16830574 The_Prisoner_(1923_film), pem_pos=15,
final_scores=-1.1844928, final_local_scores=-1.1007082,

```

pem=-2.8332133, similarity=-13.350162|
gt: 25008 Prisoner_of_war, pem_pos=13,
final_scores=-1.9050288, final_local_scores=-2.119985,
pem=-2.8332133, similarity=-43.767414]
2: 1839364 Cabinet_of_Iraq
[winner: 7515928 Iraq, pem_pos=0, final_scores=0.4624651,
final_local_scores=0.40669358, pem=-0.48469657,
similarity=18.862116|
RECALL MISS: 1839364 Cabinet_of_Iraq]
[winner: 322678 Cabinet_(file_format), pem_pos=12,
final_scores=-1.1604413, final_local_scores=-1.0410383,
pem=-3.543335, similarity=-7.7077417|
RECALL MISS: 1839364 Cabinet_of_Iraq]
[winner: 7617666 Elder_(United_Methodism), pem_pos=3,
final_scores=-1.4819678, final_local_scores=-1.4536662,
pem=-2.0384626, similarity=-28.205107|
RECALL MISS: 1839364 Cabinet_of_Iraq]
3: 601205 Missing_in_action no overlap with any filtered span
1: 14653 Iran
[pem_pos=0, final_scores=0.5812714,
final_local_scores=0.60217774, pem=-0.13856095,
similarity=22.813414]
2: 8550616 National_Iraqi_News_Agency
[pem_pos=0, final_scores=0.36473727,
final_local_scores=0.5519684, pem=0.0, similarity=20.56155]
3: 202988 Muhammad_Saeed_al-Sahhaf
[pem_pos=0, final_scores=0.31160712,
final_local_scores=0.3807354, pem=0.0, similarity=15.451616]
4: 14889 Iran-Iraq_War
[pem_pos=0, final_scores=0.26523244,
final_local_scores=0.23264068, pem=0.0, similarity=11.032175]
5: 5023832 International_Committee_of_the_Red_Cross
[pem_pos=0, final_scores=0.40295583,
final_local_scores=0.4201647, pem=0.0, similarity=16.628265]
6: 4492 Baghdad
[pem_pos=0, final_scores=0.3154046,
final_local_scores=0.2666437, pem=0.0, similarity=12.046893]
7: 18935732 Associated_Press
[pem_pos=0, final_scores=0.3017949,
final_local_scores=0.34600297, pem=0.0, similarity=14.415132]
1: 14653 Iran
[pem_pos=0, final_scores=0.44126797,
final_local_scores=0.44924527, pem=-0.57066226,
similarity=20.599438]
2: 7515928 Iraq
[pem_pos=0, final_scores=0.49947792,
final_local_scores=0.44712448, pem=-0.48469657,
similarity=20.068655]
3: 7515928 Iraq
[pem_pos=0, final_scores=0.4624651,
final_local_scores=0.40669358, pem=-0.48469657,
similarity=18.862116]

4: 14653 Iran
[pem_pos=0, final_scores=0.46500188,
final_local_scores=0.47517097, pem=-0.57066226,
similarity=21.373112]
5: 7515928 Iraq
[pem_pos=0, final_scores=0.6427364,
final_local_scores=0.6036125, pem=-0.25379938,
similarity=23.482914]
6: 7515928 Iraq
[pem_pos=0, final_scores=0.5599202,
final_local_scores=0.5131484, pem=-0.48469657,
similarity=22.03894]
7: 25008 Prisoner_of_war
[pem_pos=0, final_scores=0.307387,
final_local_scores=0.30451128, pem=0.0, similarity=13.176937]
8: 25008 Prisoner_of_war
[pem_pos=0, final_scores=0.15518592,
final_local_scores=0.13825476, pem=0.0, similarity=8.21551]
9: 7515928 Iraq
[pem_pos=0, final_scores=0.52886134,
final_local_scores=0.47922134, pem=-0.25379938,
similarity=19.770834]
10: 14653 Iran
[pem_pos=0, final_scores=0.42760873,
final_local_scores=0.43432462, pem=-0.57066226,
similarity=20.154175]
11: 25008 Prisoner_of_war
[pem_pos=0, final_scores=0.21691777,
final_local_scores=0.2056874, pem=0.0, similarity=10.227836]
12: 5023832 International_Committee_of_the_Red_Cross
[pem_pos=0, final_scores=0.23721771,
final_local_scores=0.23912099, pem=0.0, similarity=11.22556]
13: 25008 Prisoner_of_war
[pem_pos=0, final_scores=0.23875214,
final_local_scores=0.22953814, pem=0.0, similarity=10.939589]
14: 14653 Iran
[pem_pos=0, final_scores=0.58560234,
final_local_scores=0.6069087, pem=-0.13856095,
similarity=22.954596]
15: 7515928 Iraq
[pem_pos=0, final_scores=0.35211015,
final_local_scores=0.28614753, pem=-0.48469657,
similarity=15.264782]
16: 7515928 Iraq
[pem_pos=0, final_scores=0.5796427,
final_local_scores=0.53469235, pem=-0.25379938,
similarity=21.426199]
17: 25008 Prisoner_of_war
[pem_pos=0, final_scores=0.10342346,
final_local_scores=0.0817121, pem=0.0, similarity=6.528165]
18: 14653 Iran
[pem_pos=1, final_scores=0.20746844,

```
final_local_scores=0.19533867, pem=-1.0731714,  
similarity=15.75508]  
19: 14889 Iran-Iraq_War  
    [pem_pos=0, final_scores=0.23613505,  
final_local_scores=0.20085621, pem=0.0, similarity=10.083663]
```

Sample: b'APW19990120_0179_htm'

[1 Sotheby 1] 's Announces Internet Site [2 NEW YORK 2] -
LRB- AP -RRB- -- Home shoppers can buy airline tickets , 1 Be
anie Babies and garden tools via computer . Why not a \$ 9
million [1 1 Rembrandt 1] painting ? [2 2 Sotheby 2] 's , the
[3 3 New York 3] 2 auction house , said Tuesday that it will
spend \$ 25 million to launch sothebys.com , an [3 Internet 3]
site where the public will be able to bid on 3 antiques ,
art , jewelry , 4 sports memorabilia and other 5 collectibles
. The value of items on the virtual auction block initially
will be capped at about \$ 10,000 . More expensive items will
be posted when imaging technology improves enough to give
clear views of items from multiple angles . `` There 's no
point in having a fuzzy image on a \$ 100,000 item , '' said
Susan Solomon , head of the new venture . [4 Sotheby 4] 's
first major 6 online auction , planned for July , will
feature 7 baseball memorabilia . While online auction sites
have become extremely popular , consumer groups say the sites
have produced more consumer complaints than any other
Internet category because of shoddy merchandise and other
fraud . [5 Sotheby 5] 's said it would only sell merchandise
from the auction house itself or reputable art dealers . And
bids will only be accepted from customers who have registered
their 8 charge card numbers in advance . Officials at 9 Chris
tie 's , another leading [6 New York 6] auction house , said
they were also considering an online auction site .

1: 4254144 Rembrandt

2: 541267 Sotheby's

3: 645042 New_York_City

1: 312133 Beanie_Babies

[winner: 312133 Beanie_Babies, pem_pos=0,
final_scores=-0.11216013, final_local_scores=-0.048273116,
pem=0.0, similarity=2.6491468]

gt: 312133 Beanie_Babies, pem_pos=0,
final_scores=-0.11216013, final_local_scores=-0.048273116,
pem=0.0, similarity=2.6491468]

2: 62912 Auction

[winner: 16752663 House,_North_Carolina, pem_pos=16,
final_scores=-1.4509383, final_local_scores=-1.3748703,
pem=-4.8077354, similarity=-10.793949]

RECALL MISS: 62912 Auction]

[winner: 62912 Auction, pem_pos=0,
final_scores=-1.5906689, final_local_scores=-1.7855165,
pem=-0.038607303, similarity=-48.98371]

gt: 62912 Auction, pem_pos=0, final_scores=-1.5906689,
final_local_scores=-1.7855165, pem=-0.038607303,

```
similarity=-48.98371 ]
3: 346477 Antique
  [winner: 346477 Antique, pem_pos=0,
  final_scores=-1.3634036, final_local_scores=-1.4928458,
  pem=-0.0009794262, similarity=-40.45446|
  gt: 346477 Antique, pem_pos=0, final_scores=-1.3634036,
  final_local_scores=-1.4928458, pem=-0.0009794262,
  similarity=-40.45446 ]
4: 242544 Sports_memorabilia
  [winner: 242544 Sports_memorabilia, pem_pos=0,
  final_scores=-1.1755145, final_local_scores=-1.2234333,
  pem=-0.11211732, similarity=-31.810265|
  gt: 242544 Sports_memorabilia, pem_pos=0,
  final_scores=-1.1755145, final_local_scores=-1.2234333,
  pem=-0.11211732, similarity=-31.810265 ]
5: 60659 Collectable
  [winner: 60659 Collectable, pem_pos=0,
  final_scores=-1.4847579, final_local_scores=-1.6157756,
  pem=-0.009875441, similarity=-44.07455|
  gt: 60659 Collectable, pem_pos=0,
  final_scores=-1.4847579, final_local_scores=-1.6157756,
  pem=-0.009875441, similarity=-44.07455 ]
6: 840170 Online_auction
  [winner: 840170 Online_auction, pem_pos=0,
  final_scores=-1.1904131, final_local_scores=-1.2729287,
  pem=0.0, similarity=-33.89702|
  gt: 840170 Online_auction, pem_pos=0,
  final_scores=-1.1904131, final_local_scores=-1.2729287,
  pem=0.0, similarity=-33.89702 ]
7: 3850 Baseball
  [winner: 3850 Baseball, pem_pos=0,
  final_scores=-1.8571072, final_local_scores=-1.9788835,
  pem=-0.069975354, similarity=-54.58358|
  gt: 3850 Baseball, pem_pos=0, final_scores=-1.8571072,
  final_local_scores=-1.9788835, pem=-0.069975354,
  similarity=-54.58358 ]
8: 311287 Charge_card
  [winner: 311287 Charge_card, pem_pos=0,
  final_scores=-1.2548131, final_local_scores=-1.2743757,
  pem=0.0, similarity=-33.940197|
  gt: 311287 Charge_card, pem_pos=0,
  final_scores=-1.2548131, final_local_scores=-1.2743757,
  pem=0.0, similarity=-33.940197 ]
9: 155688 Christie's
  [winner: 67118 Linford_Christie, pem_pos=16,
  final_scores=-0.14364684, final_local_scores=-0.083161056,
  pem=-3.8537984, similarity=22.565577|
  RECALL MISS: 155688 Christie's]
1: 4254144 Rembrandt
  [pem_pos=0, final_scores=0.092013076,
  final_local_scores=0.1091184, pem=0.0, similarity=7.3460236]
2: 541267 Sotheby's
```

[pem_pos=0, final_scores=0.2410795,
 final_local_scores=0.18278551, pem=-0.014445427,
 similarity=9.622953]
 3: 645042 New_York_City
 [pem_pos=1, final_scores=0.43191612,
 final_local_scores=0.41450745, pem=-1.7906687,
 similarity=26.197376]
 1: 541267 Sotheby's
 [pem_pos=0, final_scores=0.16824935,
 final_local_scores=0.10322967, pem=-0.014445427,
 similarity=7.2488484]
 2: 645042 New_York_City
 [pem_pos=1, final_scores=0.19818519,
 final_local_scores=0.15919197, pem=-1.7906687,
 similarity=18.578253]
 3: 33139 World_Wide_Web
 [pem_pos=1, final_scores=0.073190615,
 final_local_scores=0.099324346, pem=-3.3911073,
 similarity=25.495117]
 4: 541267 Sotheby's
 [pem_pos=0, final_scores=0.23094098,
 final_local_scores=0.17171076, pem=-0.014445427,
 similarity=9.29246]
 5: 541267 Sotheby's
 [pem_pos=0, final_scores=0.21850674,
 final_local_scores=0.1581282, pem=-0.014445427,
 similarity=8.88713]
 6: 645042 New_York_City
 [pem_pos=1, final_scores=0.42531723,
 final_local_scores=0.40729922, pem=-1.7906687,
 similarity=25.98227]

Sample: b'APW19980713_0449_htm'

[1 Lufthansa 1] , [2 1 German 2] railway pledge closer
 cooperation Seeking to shift some domestic travel from 2 plan
 es to 3 trains , [1 1 Germany 1] 's flag carrier , [2 2 Lufth
 ansa 2] , and the [3 German 3] railway signed a letter of
 intent Monday to work on coordinating services and
 schedules . The goal is to make it easier to travel by train
 to [4 Lufthansa 4] 's 4 main hub in [3 3 Frankfurt 3] ,
 instead of flying there . Starting in 2001 , [5 Lufthansa 5]
 boss Juergen Weber hopes to eliminate such short hops _
 between [6 Frankfurt 6] and [4 4 Cologne 4] , for example _
 thus freeing up takeoff and 5 landing slots for more
 lucrative international flights . As part of the [7 Lufthansa
 7] - [5 5 Deutsche Bahn 5] deal , passengers could check
 their luggage through at the train station for the [8
 Lufthansa 8] flight they will catch at [9 Frankfurt 9] . The
 service is already in place at [10 Saarbruecken 10] ,
 although travelers must drop their bags off the day before .
 Starting in September , the service will be extended to five
 other cities : [11 Duesseldorf 11] , [12 Cologne 12] , [6 6 Bo

nn 6] , [7 7 Wuerzburg 7] and [13 Nuremberg 13] . Passengers flying on [14 Lufthansa 14] partners 6 United and 7 SAS would also be able to use the service , he said . The airline and railway said they 'd form a working group to resolve such problems as security checks on the luggage and making sure the bags get to the airport from the train station on time . [15 Deutsche Bahn 15] chief [16 Johannes Ludewig 16] said his railway also would work to coordinate schedules better with [1 7 Lufthansa 17] flights to make connections easier , and allow for combined train and plane tickets . -LRB- pfg-aet - RRB-

1: 11867 Germany
2: 97733 Lufthansa
3: 10992 Frankfurt
4: 6187 Cologne
5: 172065 Deutsche_Bahn
6: 3295 Bonn
7: 30876585 Würzburg

1: 186748 Rail_transport_in_Germany

[winner: 11867 Germany, pem_pos=1,
final_scores=0.73310393, final_local_scores=0.7368024,
pem=-1.0465823, similarity=31.768843]

RECALL MISS: 186748 Rail_transport_in_Germany]

[winner: 19940036 Yarck_railway_station, pem_pos=14,
final_scores=-1.6527835, final_local_scores=-1.6629112,
pem=-4.538745, similarity=-20.852474]

RECALL MISS: 186748 Rail_transport_in_Germany]

2: 46770 Fixed-wing_aircraft

[winner: 452991 Plane_(tool), pem_pos=2,
final_scores=-1.2332313, final_local_scores=-1.2149875,
pem=-1.2324673, similarity=-25.465586]

gt: 46770 Fixed-wing_aircraft, pem_pos=1,
final_scores=-1.3128974, final_local_scores=-1.3893421,
pem=-1.0590252, similarity=-31.61188]

3: 30598 Train

[winner: 20106448 Bombardier_MOVIA_C951, pem_pos=12,
final_scores=-1.1913999, final_local_scores=-1.1660783,
pem=-2.7080503, similarity=-15.981593]

gt: 30598 Train, pem_pos=0, final_scores=-1.3581364,
final_local_scores=-1.4764891, pem=-0.1991949,
similarity=-38.888416]

4: 1726928 Airline_hub

[winner: 5924988 Main_(CTA_station), pem_pos=8,
final_scores=-1.0252402, final_local_scores=-0.9702551,
pem=-3.1742558, similarity=-7.6025357]

RECALL MISS: 1726928 Airline_hub]

[winner: 5283280 Worldport_(UPS_air_hub), pem_pos=1,
final_scores=-1.116622, final_local_scores=-1.122817,
pem=-0.69264424, similarity=-25.650679]

gt: 1726928 Airline_hub, pem_pos=2,
final_scores=-1.4239444, final_local_scores=-1.5363779,
pem=-0.7897157, similarity=-37.464268]

5: 3189610 Landing_slot
[winner: 33129250 Takeoff_and_landing, pem_pos=0,
final_scores=-1.0441512, final_local_scores=-0.99559313,
pem=0.0, similarity=-25.620771]
RECALL MISS: 3189610 Landing_slot]
[winner: 94223 Muzzle_brake, pem_pos=2,
final_scores=-1.286227, final_local_scores=-1.2512574,
pem=-1.0986123, similarity=-27.27588]
gt: 3189610 Landing_slot, pem_pos=3,
final_scores=-1.8663638, final_local_scores=-1.9606826,
pem=-4.3719764, similarity=-30.645477]
6: 32307 United_Airlines
[winner: 32307 United_Airlines, pem_pos=0,
final_scores=-0.01459305, final_local_scores=-0.016265363,
pem=-1.5807481, similarity=12.200676]
gt: 32307 United_Airlines, pem_pos=0,
final_scores=-0.01459305, final_local_scores=-0.016265363,
pem=-1.5807481, similarity=12.200676]
7: 162868 Scandinavian_Airlines
[winner: 162868 Scandinavian_Airlines, pem_pos=2,
final_scores=-0.085661665, final_local_scores=-0.11068997,
pem=-2.3693376, similarity=13.6713295]
gt: 162868 Scandinavian_Airlines, pem_pos=2,
final_scores=-0.085661665, final_local_scores=-0.11068997,
pem=-2.3693376, similarity=13.6713295]
1: 11867 Germany
[pem_pos=0, final_scores=0.8036038,
final_local_scores=0.8138128, pem=-0.15529114,
similarity=29.220007]
2: 97733 Lufthansa
[pem_pos=0, final_scores=0.32067317,
final_local_scores=0.2879837, pem=0.0, similarity=12.683721]
3: 10992 Frankfurt
[pem_pos=0, final_scores=0.4676153,
final_local_scores=0.42091995, pem=-0.07062824,
similarity=17.034893]
4: 6187 Cologne
[pem_pos=0, final_scores=0.49876195,
final_local_scores=0.46995533, pem=-0.08239529,
similarity=18.562195]
5: 172065 Deutsche_Bahn
[pem_pos=0, final_scores=0.34127444,
final_local_scores=0.34524822, pem=0.0, similarity=14.392609]
6: 3295 Bonn
[pem_pos=0, final_scores=0.510595,
final_local_scores=0.49468374, pem=0.0, similarity=18.852062]
7: 30876585 Würzburg
[pem_pos=0, final_scores=0.31904155,
final_local_scores=0.3082512, pem=0.0, similarity=13.288544]
1: 97733 Lufthansa
[pem_pos=0, final_scores=0.261347,
final_local_scores=0.2231788, pem=0.0, similarity=10.749814]

2: 11867 Germany
[pem_pos=1, final_scores=0.73310393,
final_local_scores=0.7368024, pem=-1.0465823,
similarity=31.768843]
3: 11867 Germany
[pem_pos=1, final_scores=0.6767276,
final_local_scores=0.67521983, pem=-1.0465823,
similarity=29.931097]
4: 97733 Lufthansa
[pem_pos=0, final_scores=0.31825614,
final_local_scores=0.28534344, pem=0.0, similarity=12.60493]
5: 97733 Lufthansa
[pem_pos=0, final_scores=0.31555986,
final_local_scores=0.28239816, pem=0.0, similarity=12.517037]
6: 10992 Frankfurt
[pem_pos=0, final_scores=0.47880286,
final_local_scores=0.43314064, pem=-0.07062824,
similarity=17.399582]
7: 97733 Lufthansa
[pem_pos=0, final_scores=0.22816406,
final_local_scores=0.18693155, pem=0.0, similarity=9.668123]
8: 97733 Lufthansa
[pem_pos=0, final_scores=0.30277807,
final_local_scores=0.26843598, pem=0.0, similarity=12.100378]
9: 10992 Frankfurt
[pem_pos=0, final_scores=0.41969627,
final_local_scores=0.36857563, pem=-0.07062824,
similarity=15.472833]
10: 140379 Saarbrücken
[pem_pos=0, final_scores=0.36939555,
final_local_scores=0.36015502, pem=0.0, similarity=14.837457]
11: 65711 Düsseldorf
[pem_pos=0, final_scores=0.437082,
final_local_scores=0.40668553, pem=0.0, similarity=16.22602]
12: 6187 Cologne
[pem_pos=0, final_scores=0.54139984,
final_local_scores=0.5165307, pem=-0.08239529,
similarity=19.952097]
13: 21287 Nuremberg
[pem_pos=0, final_scores=0.2758599,
final_local_scores=0.23768198, pem=-0.14965801,
similarity=11.996481]
14: 97733 Lufthansa
[pem_pos=0, final_scores=0.34951735,
final_local_scores=0.31949157, pem=0.0, similarity=13.623979]
15: 172065 Deutsche_Bahn
[pem_pos=0, final_scores=0.42029124,
final_local_scores=0.43156213, pem=0.0, similarity=16.968388]
16: 28758695 Johannes_Ludewig
[pem_pos=0, final_scores=0.10188423,
final_local_scores=0.21774295, pem=0.0, similarity=10.587597]
17: 97733 Lufthansa

[pem_pos=0, final_scores=0.34297562,
final_local_scores=0.3132136, pem=0.0, similarity=13.436631]

Sample: b'APW19981109_0152_htm'

[1 Iran 1] to complain about [1 1 U.S. 1] - radio at the [2 Hague 2] tribunal [2 2 Iran 2] will protest to the [3 International Court of Justice 3] at 3 the [3 Hague 3] and other global bodies about the U.S.-funded [4 4 Radio Free Europe 4] , the Iran Daily reported Monday . It quoted Foreign Minister [5 5 Kamal Kharrazi 5] as saying the radio `` was set up to interfere in [4 Iran 4] 's internal affairs . '' It did not say when the complaints will be filed . The [5 English-language 5] daily also did not say which organizations [6 Iran 6] will approach besides the [6 6 ICJ 6] in 7 the [7 Netherlands 7] . On Tuesday , [7 Iran 7] recalled its ambassador from [8 8 Prague 8] to protest the [9 9 Czech 9] government 's assistance to the radio and said it would reduce the level of its economic and political cooperation with [8 Prague 8] . The American-sponsored station also began broadcasts directed to [10 10 Iraq 10] on Friday . In a trial period of several weeks , the station will broadcast one 30-minute program a day to [9 Iran 9] and [10 Iraq 10] . The [11 11 Farsi 11] - language service to [11 Iran 11] was approved by the [12 1 Czech 12] government in August . [13 Radio Free Europe 13] began transmitting from [12 12 Munich 12] , [14 Germany 14] , in 1951 , spreading uncensored news to [13 13 Soviet 13] - controlled countries behind the [14 14 Iron Curtain 14] during the [15 15 Cold War 15] between the West and the [15 2 Communist East 15] . It moved its headquarters to [16 Prague 16] in 1995 following the collapse of communism six years earlier .

1: 3434750 United_States

2: 14653 Iran

3: 30269 The_Hague

4: 184408 Radio_Free_Europe/Radio_Liberty

5: 186801 Kamal_Kharrazi

6: 14918 International_Court_of_Justice

7: 21148 Netherlands

8: 23844 Prague

9: 5321 Czech_Republic

10: 7515928 Iraq

11: 11600 Persian_language

12: 19058 Munich

13: 26779 Soviet_Union

14: 18953051 Iron_Curtain

15: 325329 Cold_War

1: 5848 Politics_of_the_Czech_Republic

[winner: 5321 Czech_Republic, pem_pos=1,
final_scores=0.5151977, final_local_scores=0.5129448,
pem=-1.208016, similarity=25.966389]

RECALL MISS: 5848 Politics_of_the_Czech_Republic]

[winner: 12229 Government, pem_pos=0,

final_scores=-1.574808, final_local_scores=-1.7232251,
pem=-0.3899827, similarity=-45.213974]
RECALL MISS: 5848 Politics_of_the_Czech_Republic]
2: 9209651 Communism
[winner: 97477 Eastern_Bloc, pem_pos=0,
final_scores=0.17980649, final_local_scores=0.16747248,
pem=-0.6931472, similarity=12.856868]
RECALL MISS: 9209651 Communism]
1: 3434750 United_States
[pem_pos=0, final_scores=0.44076997,
final_local_scores=0.42401952, pem=-0.14075686,
similarity=17.50876]
2: 14653 Iran
[pem_pos=0, final_scores=0.6429369,
final_local_scores=0.68918073, pem=-0.13856095,
similarity=25.409758]
3: 30269 The_Hague
[pem_pos=0, final_scores=0.3317626,
final_local_scores=0.32751274, pem=-0.55011845,
similarity=16.854977]
4: 184408 Radio_Free_Europe/Radio_Liberty
[pem_pos=0, final_scores=0.4157607,
final_local_scores=0.43382668, pem=0.0, similarity=17.035967]
5: 186801 Kamal_Kharazi
[pem_pos=0, final_scores=0.39243,
final_local_scores=0.49779928, pem=0.0, similarity=18.945036]
6: 14918 International_Court_of_Justice
[pem_pos=0, final_scores=0.17617036,
final_local_scores=0.19402671, pem=0.0, similarity=9.879857]
7: 21148 Netherlands
[pem_pos=0, final_scores=0.45595092,
final_local_scores=0.45467782, pem=-0.24940795,
similarity=19.014526]
8: 23844 Prague
[pem_pos=0, final_scores=0.43798065,
final_local_scores=0.44904494, pem=-0.05392049,
similarity=17.783337]
9: 5321 Czech_Republic
[pem_pos=1, final_scores=0.4872324,
final_local_scores=0.48239696, pem=-1.208016,
similarity=25.05478]
10: 7515928 Iraq
[pem_pos=0, final_scores=0.5400079,
final_local_scores=0.53700554, pem=-0.25379938,
similarity=21.49523]
11: 11600 Persian_language
[pem_pos=0, final_scores=0.055005714,
final_local_scores=0.045873657, pem=-0.17037486,
similarity=6.385199]
12: 19058 Munich
[pem_pos=0, final_scores=0.6128766,
final_local_scores=0.63576174, pem=-0.119245164,

similarity=23.710585]
13: 26779 Soviet_Union
[pem_pos=0, final_scores=0.30881554,
final_local_scores=0.28055882, pem=-0.097171314,
similarity=12.990581]
14: 18953051 Iron_Curtain
[pem_pos=0, final_scores=0.21410687,
final_local_scores=0.20131725, pem=0.0, similarity=10.097422]
15: 325329 Cold_War
[pem_pos=0, final_scores=0.24650548,
final_local_scores=0.24812955, pem=0.0, similarity=11.494393]
1: 14653 Iran
[pem_pos=0, final_scores=0.52583486,
final_local_scores=0.5612646, pem=-0.13856095,
similarity=21.592484]
2: 30269 The_Hague
[pem_pos=0, final_scores=0.23528753,
final_local_scores=0.22212842, pem=-0.55011845,
similarity=13.710098]
3: 14918 International_Court_of_Justice
[pem_pos=0, final_scores=0.26067019,
final_local_scores=0.2863299, pem=0.0, similarity=12.634368]
4: 14653 Iran
[pem_pos=0, final_scores=0.6378358,
final_local_scores=0.6836086, pem=-0.13856095,
similarity=25.243473]
5: 8569916 English_language
[pem_pos=0, final_scores=0.22738115,
final_local_scores=0.24040696, pem=0.0, similarity=11.263936]
6: 14653 Iran
[pem_pos=0, final_scores=0.6453698,
final_local_scores=0.6918384, pem=-0.13856095,
similarity=25.489067]
7: 14653 Iran
[pem_pos=0, final_scores=0.5493402,
final_local_scores=0.58694065, pem=-0.13856095,
similarity=22.35871]
8: 23844 Prague
[pem_pos=0, final_scores=0.076481,
final_local_scores=0.054161504, pem=-0.05392049,
similarity=5.9992285]
9: 14653 Iran
[pem_pos=0, final_scores=0.53987664,
final_local_scores=0.57660306, pem=-0.13856095,
similarity=22.050215]
10: 7515928 Iraq
[pem_pos=0, final_scores=0.56366724,
final_local_scores=0.5628498, pem=-0.25379938,
similarity=22.266476]
11: 14653 Iran
[pem_pos=0, final_scores=0.58487433,
final_local_scores=0.62575626, pem=-0.13856095,

similarity=23.517044]
 12: 5321 Czech_Republic
 [pem_pos=1, final_scores=0.5151977,
 final_local_scores=0.5129448, pem=-1.208016,
 similarity=25.966389]
 13: 184408 Radio_Free_Europe/Radio_Liberty
 [pem_pos=0, final_scores=0.45507568,
 final_local_scores=0.4767723, pem=0.0, similarity=18.317549]
 14: 11867 Germany
 [pem_pos=0, final_scores=0.8183802,
 final_local_scores=0.83025587, pem=-0.15529114,
 similarity=29.710701]
 15: 97477 Eastern_Bloc
 [pem_pos=0, final_scores=0.17980649,
 final_local_scores=0.16747248, pem=-0.6931472,
 similarity=12.856868]
 16: 23844 Prague
 [pem_pos=0, final_scores=0.3664539,
 final_local_scores=0.3709128, pem=-0.05392049,
 similarity=15.451717]

Sample: b'APW19990827_0137_htm'

Bank Fires Seven for E-Mail Misuse CHARLOTTE -LRB- AP -RRB-
 -- 1 First Union Corp. says it has fired seven employees
 within the last week for sending 2 pornographic and other
 inappropriate 3 e-mail . First Union , the nation 's sixth-
 largest bank , confirmed the firings Friday , but company
 spokeswoman Sandy Deem would not release more details . The
 bank `` will not tolerate conduct that interferes with
 another person 's work performance or creates an
 intimidating , offensive or 4 hostile environment , '' Deem
 said . 5 The Charlotte Observer reported Friday that
 employees at the [1 University Research Park 1] office
 circulated e-mail that included videos of people having sex
 and other material of a sexual nature . The company stumbled
 on the e-mail when the volume of messages became so heavy it
 slowed down a 6 server . One employee fired within the past
 week from the bank 's 7 home-equity division was called into
 an office , shown a stack of e-mail and fired on the spot ,
 an employee told the newspaper . First Union , a [1 1 Fortune
 100 1] corporation , joins a number of companies taking swift
 and severe action for violations of company e-mail policy .
 Companies are seeking to protect themselves from 8 legal clai
 ms that their work environment is hostile . An increasing
 number of lawsuits alleging 9 sexual harassment have based
 claims in part on e-mails taken from company computer
 systems .

1: 276447 Fortune_500

1: 2469065 First_Union

[winner: 20918077

National_Register_of_Historic_Places_listings_in_Union_County
 ,_Pennsylvania, pem_pos=6, final_scores=-1.4870437,

final_local_scores=-1.4861917, pem=-4.1654887,
 similarity=-17.60864|
 RECALL MISS: 2469065 First_Union]
 [winner: 31653
 First_Amendment_to_the_United_States_Constitution, pem_pos=1,
 final_scores=-1.6187567, final_local_scores=-1.7112691,
 pem=-2.36835, similarity=-34.09852|
 RECALL MISS: 2469065 First_Union]
 2: 5710507 Pornography
 [winner: 238253 Pornographic_film, pem_pos=0,
 final_scores=-1.5550644, final_local_scores=-1.6254108,
 pem=0.0, similarity=-44.415787|
 gt: 5710507 Pornography, pem_pos=1,
 final_scores=-1.5746654, final_local_scores=-1.6548817,
 pem=-0.1339556, similarity=-44.566784]
 3: 9738 Email
 [winner: 472401 Header_(computing), pem_pos=0,
 final_scores=-1.1178211, final_local_scores=-1.053346,
 pem=-0.15179636, similarity=-26.518742|
 gt: 9738 Email, pem_pos=3, final_scores=-1.8136023,
 final_local_scores=-1.9054823, pem=-2.2651007,
 similarity=-40.45571]
 4: 1471574 Hostile_environment_sexual_harassment
 [winner: 1471574 Hostile_environment_sexual_harassment,
 pem_pos=2, final_scores=-1.1876172,
 final_local_scores=-1.1201978, pem=-1.4271164,
 similarity=-21.578339|
 gt: 1471574 Hostile_environment_sexual_harassment,
 pem_pos=2, final_scores=-1.1876172,
 final_local_scores=-1.1201978, pem=-1.4271164,
 similarity=-21.578339]
 5: 2349275 The_Charlotte_Observer
 [winner: 2349275 The_Charlotte_Observer, pem_pos=0,
 final_scores=-0.00397557, final_local_scores=0.06175278,
 pem=0.0, similarity=5.932539|
 gt: 2349275 The_Charlotte_Observer, pem_pos=0,
 final_scores=-0.00397557, final_local_scores=0.06175278,
 pem=0.0, similarity=5.932539]
 6: 42116 Server_(computing)
 [winner: 42116 Server_(computing), pem_pos=0,
 final_scores=-0.9930034, final_local_scores=-0.9954402,
 pem=0.0, similarity=-25.616207|
 gt: 42116 Server_(computing), pem_pos=0,
 final_scores=-0.9930034, final_local_scores=-0.9954402,
 pem=0.0, similarity=-25.616207]
 7: 2662518 Home_equity
 [winner: 2662518 Home_equity, pem_pos=0,
 final_scores=-1.1381774, final_local_scores=-1.1225315,
 pem=0.0, similarity=-29.408867|
 gt: 2662518 Home_equity, pem_pos=0,
 final_scores=-1.1381774, final_local_scores=-1.1225315,
 pem=0.0, similarity=-29.408867]

8: 3221417 Cause_of_action
 [winner: 872847 SC0-Linux_controversies, pem_pos=0,
 final_scores=-1.2136945, final_local_scores=-1.2407367,
 pem=0.0, similarity=-32.936344]
 RECALL MISS: 3221417 Cause_of_action]

9: 236331 Sexual_harassment
 [winner: 9022677 Sexual_obsessions, pem_pos=3,
 final_scores=-1.2346692, final_local_scores=-1.1565294,
 pem=-1.8493533, similarity=-20.36636]
 RECALL MISS: 236331 Sexual_harassment]
 [winner: 22216147 Harassment, pem_pos=0,
 final_scores=-1.3512793, final_local_scores=-1.3752153,
 pem=-0.1074169, similarity=-36.365307]
 gt: 236331 Sexual_harassment, pem_pos=1,
 final_scores=-2.0425153, final_local_scores=-2.1530244,
 pem=-3.3892677, similarity=-41.729465]

1: 276447 Fortune_500
 [pem_pos=0, final_scores=0.14829563,
 final_local_scores=0.19068053, pem=0.0, similarity=9.780001]

1: 6190443 University_City_(Charlotte_neighborhood)
 [pem_pos=0, final_scores=0.08373539,
 final_local_scores=0.15840843, pem=0.0, similarity=8.8169365]

Sample: b'APW19981210_0433_htm'

[1 Syria 1] opposes any attempt to overthrow [2 Saddam Hussein 2] [1 1 Syria 1] opposes attempts by the [2 2 United States 2] and [3 3 Britain 3] to overthrow [4 4 Iraqi 4] leader [5 5 Saddam Hussein 5] , [3 Syrian 3] Vice-President [6 6 Abdul-Halim Khaddam 6] said Thursday . Such moves would ``serve [7 7 Israel 7] 's design to disintegrate the [4 1 Arab 4] and [8 8 Islamic 8] world , '' [5 Khaddam 5] said after receiving his [6 Iranian 6] counterpart , [9 9 Hassan Habibi 9] , at [7 2 Damascus 7] airport . The [8 United States 8] has granted dlrs 97 million to [9 Iraqi 9] opposition groups . [10 British 10] and [11 American 11] officials hosted meetings in [10 10 London 10] last month of 16 [12 Iraqi 12] dissident groups in a bid to promote cooperation among them . `` This is a very dangerous issue ... No state in the world has the right to interfere in the internal affairs of any other country , '' [13 Khaddam 13] said . [14 Habibi 14] said the [1 5 Iraqi 15] and [16 3 Palestinian 16] questions were of great concern to [17 Iran 17] and [18 Syria 18] : `` There should be coordination between the two countries over these issues . '' The official [19 Iranian 19] news agency , IRMA , said the [20 Iraqi 20] and [21 Palestinian 21] issues would be the focus of talks between [22 Habibi 22] and [23 Khaddam 23] at the meeting of the Syrian-Iranian Supreme Committee . The committee meets biannually to promote cooperation in foreign affairs , trade and technology . [24 Habibi 24] , who is accompanied by [25 11 Iranian 25] Foreign Minister [11 Kamal K harrazi 11] , is expected to meet [26 Syrian 26] President [12 12 Hafez Assad 12] during his three-day visit . [27 Syria 27]

is [28 Iran 28] 's closest [29 Arab 29] ally and supported
the [30 Persian 30] state during the 1980-88 [13 13 Iran-Iraq
13] war . str-bm-jbm

1: 7515849 Syria
2: 3434750 United_States
3: 31717 United_Kingdom
4: 7515928 Iraq
5: 29490 Saddam_Hussein
6: 3044825 Abdul_Halim_Khaddam
7: 9282173 Israel
8: 6037917 Islam
9: 1070742 Hassan_Habibi
10: 17867 London
11: 186801 Kamal_Kharazi
12: 70843 Hafez_al-Assad
13: 14889 Iran-Iraq_War

1: 2185 Arab_people
[winner: 159433 Arab_world, pem_pos=14,
final_scores=0.43327808, final_local_scores=0.41578853,
pem=-3.8052368, similarity=37.191143]
gt: 2185 Arab_people, pem_pos=1,
final_scores=-0.046289697, final_local_scores=-0.10669571,
pem=-2.0453336, similarity=12.0285425]
2: 4202280 Damascus_International_Airport
[winner: 8914 Damascus, pem_pos=0,
final_scores=0.30604088, final_local_scores=0.26133326,
pem=-0.120337, similarity=12.5428295]
gt: 4202280 Damascus_International_Airport, pem_pos=14,
final_scores=-0.66794956, final_local_scores=-0.67767185,
pem=-3.1456103, similarity=0.97295225]
[winner: 37575 Airport, pem_pos=0,
final_scores=-1.4976324, final_local_scores=-1.5156994,
pem=-0.121571906, similarity=-40.480656]
RECALL MISS: 4202280 Damascus_International_Airport]

3: 23267 Palestinian_people
[winner: 241405 State_of_Palestine, pem_pos=9,
final_scores=0.56226724, final_local_scores=0.5855869,
pem=-2.8508017, similarity=37.067898]
gt: 23267 Palestinian_people, pem_pos=0,
final_scores=-0.37753397, final_local_scores=-0.4729262,
pem=-0.38816506, similarity=-7.912446]

1: 7515849 Syria
[pem_pos=0, final_scores=0.62075406,
final_local_scores=0.59098864, pem=-0.26162508,
similarity=23.14875]
2: 3434750 United_States
[pem_pos=0, final_scores=0.5037217,
final_local_scores=0.5179967, pem=-0.073457584,
similarity=19.947243]
3: 31717 United_Kingdom
[pem_pos=0, final_scores=0.49917746,
final_local_scores=0.5124087, pem=-0.61114657,

similarity=22.704521]
4: 7515928 Iraq
[pem_pos=0, final_scores=0.53225064,
final_local_scores=0.4824757, pem=-0.48469657,
similarity=21.123606]
5: 29490 Saddam_Hussein
[pem_pos=0, final_scores=0.38556224,
final_local_scores=0.36855942, pem=0.0, similarity=15.088261]
6: 3044825 Abdul_Halim_Khaddam
[pem_pos=0, final_scores=0.5020225,
final_local_scores=0.5561342, pem=0.0, similarity=20.685867]
7: 9282173 Israel
[pem_pos=0, final_scores=0.5664993,
final_local_scores=0.55256826, pem=-0.21279564,
similarity=21.736668]
8: 6037917 Islam
[pem_pos=0, final_scores=0.36058706,
final_local_scores=0.3731748, pem=-0.21141872,
similarity=16.375721]
9: 1070742 Hassan_Habibi
[pem_pos=0, final_scores=0.2374631,
final_local_scores=0.30434138, pem=0.0, similarity=13.171866]
10: 17867 London
[pem_pos=0, final_scores=0.49195492,
final_local_scores=0.53402203, pem=-0.09086737,
similarity=20.520145]
11: 186801 Kamal_Kharazi
[pem_pos=0, final_scores=0.41640753,
final_local_scores=0.48311335, pem=0.0, similarity=18.506779]
12: 70843 Hafez_al-Assad
[pem_pos=0, final_scores=0.44244224,
final_local_scores=0.4468845, pem=0.0, similarity=17.425638]
13: 14889 Iran-Iraq_War
[pem_pos=0, final_scores=0.26841074,
final_local_scores=0.24120125, pem=0.0, similarity=11.28764]
1: 7515849 Syria
[pem_pos=0, final_scores=0.51522875,
final_local_scores=0.47571832, pem=-0.26162508,
similarity=19.708855]
2: 29490 Saddam_Hussein
[pem_pos=0, final_scores=0.1762409,
final_local_scores=0.13990763, pem=0.0, similarity=8.264835]
3: 7515849 Syria
[pem_pos=0, final_scores=0.59551674,
final_local_scores=0.5634207, pem=-0.60890055,
similarity=24.214607]
4: 159433 Arab_world
[pem_pos=14, final_scores=0.43327808,
final_local_scores=0.41578853, pem=-3.8052368,
similarity=37.191143]
5: 3044825 Abdul_Halim_Khaddam
[pem_pos=0, final_scores=0.16508482,

```
final_local_scores=0.18808103, pem=0.0, similarity=9.702426]
6: 14653 Iran
    [pem_pos=0, final_scores=0.40306157,
final_local_scores=0.42420286, pem=-0.57066226,
similarity=19.852121]
7: 8914 Damascus
    [pem_pos=0, final_scores=0.30604088,
final_local_scores=0.26133326, pem=-0.120337,
similarity=12.5428295]
8: 3434750 United_States
    [pem_pos=0, final_scores=0.46056467,
final_local_scores=0.47085422, pem=-0.073457584,
similarity=18.540417]
9: 7515928 Iraq
    [pem_pos=0, final_scores=0.5371298,
final_local_scores=0.4878055, pem=-0.48469657,
similarity=21.282658]
10: 31717 United_Kingdom
    [pem_pos=0, final_scores=0.42014045,
final_local_scores=0.42607278, pem=-0.48936236,
similarity=19.465803]
11: 3434750 United_States
    [pem_pos=0, final_scores=0.45047414,
final_local_scores=0.45983183, pem=-0.12994486,
similarity=18.518673]
12: 7515928 Iraq
    [pem_pos=0, final_scores=0.55813414,
final_local_scores=0.5107495, pem=-0.48469657,
similarity=21.967352]
13: 3044825 Abdul_Halim_Khaddam
    [pem_pos=0, final_scores=0.155977,
final_local_scores=0.17813209, pem=0.0, similarity=9.40553]
14: 1070742 Hassan_Habibi
    [pem_pos=0, final_scores=0.18753134,
final_local_scores=0.24924147, pem=0.0, similarity=11.5275755]
15: 7515928 Iraq
    [pem_pos=0, final_scores=0.5334187,
final_local_scores=0.48375165, pem=-0.48469657,
similarity=21.161682]
16: 241405 State_of_Palestine
    [pem_pos=9, final_scores=0.56226724,
final_local_scores=0.5855869, pem=-2.8508017,
similarity=37.067898]
17: 14653 Iran
    [pem_pos=0, final_scores=0.6373982,
final_local_scores=0.6801801, pem=-0.13856095,
similarity=25.14116]
18: 7515849 Syria
    [pem_pos=0, final_scores=0.5522795,
final_local_scores=0.5161906, pem=-0.26162508,
similarity=20.916628]
19: 14653 Iran
```

[pem_pos=0, final_scores=0.51857346,
 final_local_scores=0.5503821, pem=-0.57066226,
 similarity=23.617561]
 20: 7515928 Iraq
 [pem_pos=0, final_scores=0.54266703,
 final_local_scores=0.49385405, pem=-0.48469657,
 similarity=21.463158]
 21: 241405 State_of_Palestine
 [pem_pos=9, final_scores=0.45874566,
 final_local_scores=0.4725052, pem=-2.8508017,
 similarity=33.693314]
 22: 1070742 Hassan_Habibi
 [pem_pos=0, final_scores=0.20380189,
 final_local_scores=0.26701456, pem=0.0, similarity=12.0579605]
 23: 3044825 Abdul_Halim_Khaddam
 [pem_pos=0, final_scores=0.2072195,
 final_local_scores=0.23410675, pem=0.0, similarity=11.075925]
 24: 1070742 Hassan_Habibi
 [pem_pos=0, final_scores=0.13793127,
 final_local_scores=0.19506091, pem=0.0, similarity=9.91072]
 25: 14653 Iran
 [pem_pos=0, final_scores=0.4389091,
 final_local_scores=0.4633609, pem=-0.57066226,
 similarity=21.020676]
 26: 7515849 Syria
 [pem_pos=0, final_scores=0.62098664,
 final_local_scores=0.59124273, pem=-0.60890055,
 similarity=25.044872]
 27: 7515849 Syria
 [pem_pos=0, final_scores=0.57917887,
 final_local_scores=0.54557407, pem=-0.26162508,
 similarity=21.79349]
 28: 14653 Iran
 [pem_pos=0, final_scores=0.6273692,
 final_local_scores=0.66922486, pem=-0.13856095,
 similarity=24.814234]
 29: 159433 Arab_world
 [pem_pos=14, final_scores=0.37941593,
 final_local_scores=0.3569523, pem=-3.8052368,
 similarity=35.435352]
 30: 14653 Iran
 [pem_pos=3, final_scores=0.18766163,
 final_local_scores=0.18891117, pem=-2.251121,
 similarity=21.969145]

Sample: b'APW19981010_0354_htm'

[1 Cohen 1] : Missile defense system in [1 1 Gulf 1] `imperative` [2 U.S. 2] Defense Secretary [2 2 William Cohen 2] said Saturday it has become ``imperative`` for [3 3 U.S. 3] allies in the [3 Gulf 3] to acquire a 1 ballistic missile defense system to counter long range weapons in the hands of countries such as [4 4 Iran 4] and [5 5 Iraq 5] . [4 Cohen 4]

said the [5 United States 5] has five defense systems under research and development and `` we hope we hope we can share that technology with [6 6 Bahrain 6] and other [6 Gulf 6] countries . '' He was speaking to reporters before leaving for [7 7 Qatar 7] , his second stop in a six-nation [7 Gulf 7] tour , which includes the [8 8 United Arab Emirates 8] , [9 9 Oman 9] , [8 Kuwait 8] and [10 10 Saudi Arabia 10] . `` Because of the proliferation of missile technology globally but especially in the [9 Gulf 9] region , we believe it 's important '' to look at ways to cooperate with [10 U.S. 10] allies , said [11 Cohen 11] . `` I think it becomes imperative that countries cooperate and develop a theater missile protection system for their people and for their forces , '' said [12 Cohen 12] , citing threats from [13 Iran 13] and [14 Iraq 14] . The [15 United States 15] sees [16 Iraq 16] and [17 Iran 17] , which recently tested a long-range missile , as major threats to the [18 Gulf 18] region . [19 Cohen 19] said [20 Iraq 20] , under [11 11 U.N. 11] 2 sanctions , continues efforts to acquire missile technology that has a longer than permissible range .

1: 24761 Persian_Gulf
2: 375197 William_Cohen
3: 3434750 United_States
4: 14653 Iran
5: 7515928 Iraq
6: 18933277 Bahrain
7: 8238258 Qatar
8: 69328 United_Arab_Emirates
9: 22316 Oman
10: 349303 Saudi_Arabia
11: 31769 United_Nations

1: 350551 United_States_national_missile_defense
[winner: 36647151

Missile_defense_systems_of_various_nations, pem_pos=0,
final_scores=-0.7714145, final_local_scores=-0.81039417,
pem=0.0, similarity=-20.094065]

RECALL MISS: 350551

United_States_national_missile_defense]

[winner: 1285488 Hadès_(missile), pem_pos=3,
final_scores=-0.89776385, final_local_scores=-0.851441,
pem=-1.0986123, similarity=-15.344562]

RECALL MISS: 350551

United_States_national_missile_defense]

[winner: 584911 External_ballistics, pem_pos=2,
final_scores=-1.2801952, final_local_scores=-1.2769401,
pem=-0.58503115, similarity=-30.835236]

RECALL MISS: 350551

United_States_national_missile_defense]

2: 39411680 International_sanctions

[winner: 14469390

United_Nations_Security_Council_Resolution_282, pem_pos=4,
final_scores=-1.2639692, final_local_scores=-1.2804117,

pem=-1.7893898, similarity=-24.389349|
gt: 39411680 International_sanctions, pem_pos=0,
final_scores=-1.5650733, final_local_scores=-1.6929591,
pem=-0.629573, similarity=-43.00785]
1: 24761 Persian_Gulf
[pem_pos=4, final_scores=0.22365122,
final_local_scores=0.18619797, pem=-2.1612167,
similarity=21.399265]
2: 375197 William_Cohen
[pem_pos=0, final_scores=0.51730555,
final_local_scores=0.5548514, pem=0.0, similarity=20.647585]
3: 3434750 United_States
[pem_pos=0, final_scores=0.5212136,
final_local_scores=0.5056432, pem=-0.14075686,
similarity=19.94457]
4: 14653 Iran
[pem_pos=0, final_scores=0.5756817,
final_local_scores=0.637432, pem=-0.13856095,
similarity=23.86547]
5: 7515928 Iraq
[pem_pos=0, final_scores=0.56922233,
final_local_scores=0.5396955, pem=-0.25379938,
similarity=21.575504]
6: 18933277 Bahrain
[pem_pos=0, final_scores=0.33712035,
final_local_scores=0.29895008, pem=-0.42385748,
similarity=15.315982]
7: 8238258 Qatar
[pem_pos=0, final_scores=0.34747797,
final_local_scores=0.32312104, pem=-0.3556309,
similarity=15.666266]
8: 69328 United_Arab_Emirates
[pem_pos=0, final_scores=0.39507133,
final_local_scores=0.36806864, pem=-0.24986836,
similarity=16.432438]
9: 22316 Oman
[pem_pos=0, final_scores=0.5046053,
final_local_scores=0.49402183, pem=-0.38991782,
similarity=20.952744]
10: 349303 Saudi_Arabia
[pem_pos=0, final_scores=0.46462017,
final_local_scores=0.42626148, pem=-0.25454172,
similarity=18.194443]
11: 31769 United_Nations
[pem_pos=0, final_scores=0.49570787,
final_local_scores=0.50674695, pem=0.0, similarity=19.212053]
1: 375197 William_Cohen
[pem_pos=0, final_scores=0.28980762,
final_local_scores=0.30770755, pem=0.0, similarity=13.27232]
2: 3434750 United_States
[pem_pos=0, final_scores=0.54306,
final_local_scores=0.5295071, pem=-0.14075686,

similarity=20.656717]
3: 24761 Persian_Gulf
[pem_pos=4, final_scores=0.22777896,
final_local_scores=0.19070688, pem=-2.1612167,
similarity=21.53382]
4: 375197 William_Cohen
[pem_pos=0, final_scores=0.5345921,
final_local_scores=0.5750974, pem=0.0, similarity=21.251766]
5: 3434750 United_States
[pem_pos=0, final_scores=0.51250046,
final_local_scores=0.4961254, pem=-0.073457584,
similarity=19.29456]
6: 24761 Persian_Gulf
[pem_pos=4, final_scores=0.2017154,
final_local_scores=0.16223642, pem=-2.1612167,
similarity=20.684204]
7: 24761 Persian_Gulf
[pem_pos=4, final_scores=0.22403683,
final_local_scores=0.18661919, pem=-2.1612167,
similarity=21.411835]
8: 7515890 Kuwait
[pem_pos=0, final_scores=0.48664504,
final_local_scores=0.44884682, pem=-0.3372397,
similarity=19.31816]
9: 24761 Persian_Gulf
[pem_pos=4, final_scores=0.12241788,
final_local_scores=0.07561591, pem=-2.1612167,
similarity=18.099276]
10: 3434750 United_States
[pem_pos=0, final_scores=0.4891544,
final_local_scores=0.47062343, pem=-0.14075686,
similarity=18.899513]
11: 375197 William_Cohen
[pem_pos=0, final_scores=0.5604754,
final_local_scores=0.6020079, pem=0.0, similarity=22.05483]
12: 375197 William_Cohen
[pem_pos=0, final_scores=0.53593427,
final_local_scores=0.5752004, pem=0.0, similarity=21.254839]
13: 14653 Iran
[pem_pos=0, final_scores=0.62571985,
final_local_scores=0.69209105, pem=-0.13856095,
similarity=25.496607]
14: 7515928 Iraq
[pem_pos=0, final_scores=0.5615395,
final_local_scores=0.5313031, pem=-0.25379938,
similarity=21.325058]
15: 3434750 United_States
[pem_pos=0, final_scores=0.4719259,
final_local_scores=0.45180386, pem=-0.073457584,
similarity=17.971916]
16: 7515928 Iraq
[pem_pos=0, final_scores=0.6053818,

```

final_local_scores=0.5791942, pem=-0.25379938,
similarity=22.754223]
17: 14653 Iran
    [pem_pos=0, final_scores=0.6063012,
final_local_scores=0.6708791, pem=-0.13856095,
similarity=24.8636]
18: 24761 Persian_Gulf
    [pem_pos=4, final_scores=0.2410583,
final_local_scores=0.20521256, pem=-2.1612167,
similarity=21.966698]
19: 375197 William_Cohen
    [pem_pos=0, final_scores=0.59284407,
final_local_scores=0.64663124, pem=0.0, similarity=23.386478]
20: 7515928 Iraq
    [pem_pos=0, final_scores=0.6349983,
final_local_scores=0.6115458, pem=-0.25379938,
similarity=23.71966]

```

Sample: b'APW19980611_0774_htm'

International police seek to stymie 1 illegal art traffic king Police officers from 23 countries met in [1 1 Budapest 1] to seek more efficient ways to combat 2 illegal traffic in art treasures , a police official said Thursday . Held under the auspices of 3 Interpol , the 4 International Police Organization , the three-day closed meeting allowed the officers to make personal contact and to help their [2 2 East European 2] counterparts catch up . `` We have a five-member team working on illegal art traffic , and the group was set up only this April , while [3 3 France 3] has over 30 policemen in the same detail and with many years of experience , ' ' [1 Hungarian 1] police Col. Tamas Simon said . Illegal art traffic is a growth industry . In 1990 , the year after 5 communism collapsed , there were 702 registered cases in [2 Hungary 2] , and the total value of the stolen treasures was 71 million 6 forints -LRB- dlrs 340,000 in present terms -RRB- . In 1993 , there were in 1996 , 1,025 cases involving 430 million forints -LRB- dlrs two million -RRB- worth of stolen treasures . `` This is a special area of police work as stolen art may not appear for 30 years on the market , ' ' said Jean Pierre Jouanny , director of the [3 French 3] national bureau of Interpol . [4 Hungary 4] is largely a transit country for stolen art , with the line of traffic moving from east to west . Work is underway in cataloging the country 's art treasures , establishing a 7 data base and a 8 network of contacts among 9 art dealers , 10 auctioneers and museums . -LRB- ab/rp -RRB-

```

1: 36787 Budapest
2: 37403 Eastern_Europe
3: 5843419 France
1: 425514 Art_theft

```

```

[winner: 1307417 Illegal_Art, pem_pos=0,
final_scores=-0.995322, final_local_scores=-0.9366521,

```

pem=0.0, similarity=-23.861855|
RECALL MISS: 425514 Art_theft]
[winner: 2435642 Traffic_(2004_miniseries), pem_pos=4,
final_scores=-1.3201647, final_local_scores=-1.2787509,
pem=-2.6704595, similarity=-19.548391|
RECALL MISS: 425514 Art_theft]
2: 28645 Smuggling
[winner: 28645 Smuggling, pem_pos=0,
final_scores=-1.3677373, final_local_scores=-1.4494879,
pem=0.0, similarity=-39.165897|
gt: 28645 Smuggling, pem_pos=0, final_scores=-1.3677373,
final_local_scores=-1.4494879, pem=0.0,
similarity=-39.165897]
3: 14338129 Interpol
[winner: 14338129 Interpol, pem_pos=0,
final_scores=-0.29771715, final_local_scores=-0.28358474,
pem=-0.41894695, similarity=-2.0947225|
gt: 14338129 Interpol, pem_pos=0,
final_scores=-0.29771715, final_local_scores=-0.28358474,
pem=-0.41894695, similarity=-2.0947225]
4: 14338156 United_Nations_Police
[winner: 14338129 Interpol, pem_pos=0,
final_scores=0.01372464, final_local_scores=0.10004197,
pem=0.0, similarity=7.075165|
RECALL MISS: 14338156 United_Nations_Police]
5: 9209651 Communism
[winner: 9209651 Communism, pem_pos=0,
final_scores=-1.0413684, final_local_scores=-1.1412127,
pem=-0.07914143, similarity=-29.535963|
gt: 9209651 Communism, pem_pos=0,
final_scores=-1.0413684, final_local_scores=-1.1412127,
pem=-0.07914143, similarity=-29.535963]
6: 229795 Hungarian_forint
[winner: 3359939 Austro-Hungarian_gulden, pem_pos=1,
final_scores=-1.302099, final_local_scores=-1.3603895,
pem=-0.66860604, similarity=-32.871037|
gt: 229795 Hungarian_forint, pem_pos=0,
final_scores=-1.3371509, final_local_scores=-1.4105101,
pem=0.0, similarity=-38.002724]
7: 8377 Database
[winner: 2234333 Data_(computing), pem_pos=1,
final_scores=-1.1630232, final_local_scores=-1.112914,
pem=-0.4798127, similarity=-26.512564|
RECALL MISS: 8377 Database]
[winner: 2544227 LORAN-C_transmitter_Jan_Mayen,
pem_pos=16, final_scores=-1.3183404,
final_local_scores=-1.2344282, pem=-2.83117,
similarity=-17.351744|
RECALL MISS: 8377 Database]
8: 2903105 Entrepreneurial_network
[winner: 85024 Network_layer, pem_pos=12,
final_scores=-1.3553039, final_local_scores=-1.2997775,


```

pem=-3.1110392, similarity=-17.779926|
  gt: 2903105 Entrepreneurial_network, pem_pos=17,
final_scores=-1.5496682, final_local_scores=-1.4934506,
pem=-3.2322292, similarity=-22.900465 ]
9: 5961279 Art_dealer
  [winner: 5961279 Art_dealer, pem_pos=0,
final_scores=-1.2205292, final_local_scores=-1.2718072,
pem=0.0, similarity=-33.863552]
  gt: 5961279 Art_dealer, pem_pos=0,
final_scores=-1.2205292, final_local_scores=-1.2718072,
pem=0.0, similarity=-33.863552 ]
10: 62912 Auction
  [winner: 62912 Auction, pem_pos=0,
final_scores=-1.6160327, final_local_scores=-1.7049174,
pem=0.0, similarity=-46.788425]
  gt: 62912 Auction, pem_pos=0, final_scores=-1.6160327,
final_local_scores=-1.7049174, pem=0.0,
similarity=-46.788425 ]
1: 36787 Budapest
  [pem_pos=0, final_scores=0.62597865,
final_local_scores=0.6153102, pem=-0.030229695,
similarity=22.61619]
2: 37403 Eastern_Europe
  [pem_pos=0, final_scores=0.3160689,
final_local_scores=0.30915284, pem=0.0, similarity=13.315451]
3: 5843419 France
  [pem_pos=0, final_scores=0.5751635,
final_local_scores=0.54810834, pem=-0.16240834,
similarity=21.32956]
1: 13275 Hungary
  [pem_pos=1, final_scores=0.7002731,
final_local_scores=0.6637211, pem=-1.0319586,
similarity=29.508425]
2: 13275 Hungary
  [pem_pos=0, final_scores=0.62672526,
final_local_scores=0.58338124, pem=-0.32577527,
similarity=23.27059]
3: 5843419 France
  [pem_pos=0, final_scores=0.7346975,
final_local_scores=0.7223751, pem=-0.80223274,
similarity=30.009493]
4: 13275 Hungary
  [pem_pos=0, final_scores=0.5868786,
final_local_scores=0.5398549, pem=-0.32577527,
similarity=21.971676]

```

Sample: b'APW19981020_1367_htm'

```

[1 Shepard 1] , [2 Glenn 2] set first two milestones __ May
5 , 1961 : [1 1 Alan Shepard 1] becomes first [3 American 3]
in space . Feb. 20 , 1962 : [2 2 John Glenn 2] becomes first [
4 American 4] in 1 orbit . Jan. 27 , 1967 : [3 3 Gus Grissom 3
] , 2 Edward White II and 3 Roger Chaffee die in 4 Apollo 1 5

```

spacecraft fire on 6 launch pad . July 20 , 1969 : 7 Apollo 11 's [4 4 Neil Armstrong 4] and [5 5 Buzz Aldrin 5] land on 8 moon . July 17 , 1975 : American Apollo and Soviet 9 Soyuz spacecraft link in orbit . April 12 , 1981 : 10 Columbia soars on first 11 space shuttle flight . June 18 , 1983 : [6 6 Sally Ride 6] becomes first [5 American 5] woman in space . Jan. 28 , 1986 : 12 Challenger explodes , killing all seven on board . April 25 , 1990 : [7 7 Hubble Space Telescope 7] is released into orbit . Dec. 2 , 1993 : First [6 Hubble 6] repair mission is launched . March 14 , 1995 : [8 8 Norman Thagard 8] is first [7 American 7] to be launched on a [8 Russian 8] 13 rocket . Two days later , he becomes first [9 American 9] to visit 14 Mir . June 29 , 1995 : 15 Atlantis docks with [10 Mir 10] in first shuttle-station hookup . Sept. 26 , 1996 : [9 9 Shannon Lucid 9] returns to [11 Earth 11] after 188-day [12 Mir 12] mission , a [13 U.S. 13] space endurance record and a world record for women . Nov. 19 , 1996 : [10 10 Story Musgrave 10] , at age 61 , becomes oldest man in space . Oct. 29 , 1998 : 16 Discovery is scheduled to blast off , carrying 77-year-old [14 John Glenn 14] back into orbit and making him oldest man in space .

- 1: 63727 Alan_Shepard
- 2: 58702 John_Glenn
- 3: 36592 Gus_Grissom
- 4: 21247 Neil_Armstrong
- 5: 65777 Buzz_Aldrin
- 6: 94289 Sally_Ride
- 7: 40203 Hubble_Space_Telescope
- 8: 514369 Norman_Thagard
- 9: 501402 Shannon_Lucid
- 10: 380558 Story_Musgrave

1: 22498 Orbit

[winner: 23627083 In_Orbit_(Clark_Terry_album),
pem_pos=0, final_scores=-0.69346637,
final_local_scores=-0.58200526, pem=0.0,
similarity=-13.278486]

RECALL MISS: 22498 Orbit]

2: 36594 Edward_Higgins_White

[winner: 36594 Edward_Higgins_White, pem_pos=0,
final_scores=-0.0943259, final_local_scores=-0.11234303,
pem=0.0, similarity=0.7371728]

gt: 36594 Edward_Higgins_White, pem_pos=0,
final_scores=-0.0943259, final_local_scores=-0.11234303,
pem=0.0, similarity=0.7371728]

3: 36595 Roger_B._Chaffee

[winner: 36595 Roger_B._Chaffee, pem_pos=0,
final_scores=-0.042892963, final_local_scores=-0.040819332,
pem=0.0, similarity=2.8715827]

gt: 36595 Roger_B._Chaffee, pem_pos=0,
final_scores=-0.042892963, final_local_scores=-0.040819332,
pem=0.0, similarity=2.8715827]

4: 1965 Apollo_1

[winner: 1965 Apollo_1, pem_pos=0,
final_scores=-0.069572315, final_local_scores=-0.08531821,
pem=0.0, similarity=1.5436474|
gt: 1965 Apollo_1, pem_pos=0, final_scores=-0.069572315,
final_local_scores=-0.08531821, pem=0.0,
similarity=1.5436474]
5: 37910 Spacecraft
[winner: 37910 Spacecraft, pem_pos=0,
final_scores=-1.0411235, final_local_scores=-1.1688085,
pem=-0.13564423, similarity=-30.052212|
gt: 37910 Spacecraft, pem_pos=0,
final_scores=-1.0411235, final_local_scores=-1.1688085,
pem=-0.13564423, similarity=-30.052212]
6: 316577 Launch_pad
[winner: 5124085 Drill_floor, pem_pos=3,
final_scores=-1.0277048, final_local_scores=-0.96963555,
pem=-2.1036558, similarity=-13.406139|
gt: 316577 Launch_pad, pem_pos=10,
final_scores=-1.5856664, final_local_scores=-1.6939349,
pem=-3.95604, similarity=-24.947126]
[winner: 5102421 Land_Launch, pem_pos=3,
final_scores=-1.155859, final_local_scores=-1.1598237,
pem=-1.3862944, similarity=-22.982857|
RECALL MISS: 316577 Launch_pad]
7: 662 Apollo_11
[winner: 594 Apollo, pem_pos=0, final_scores=-0.23525368,
final_local_scores=-0.17430694, pem=-0.49267578,
similarity=1.5672934|
gt: 662 Apollo_11, pem_pos=16, final_scores=-0.754247,
final_local_scores=-0.8902609, pem=-4.087068,
similarity=-0.25133598]
[winner: 19452 May_11, pem_pos=2,
final_scores=-0.5767282, final_local_scores=-0.3180565,
pem=-3.271949, similarity=12.391641|
RECALL MISS: 662 Apollo_11]
8: 19331 Moon
[winner: 19331 Moon, pem_pos=0, final_scores=-0.5263862,
final_local_scores=-0.59796, pem=-0.23889688,
similarity=-12.455447|
gt: 19331 Moon, pem_pos=0, final_scores=-0.5263862,
final_local_scores=-0.59796, pem=-0.23889688,
similarity=-12.455447]
9: 178182 Soyuz_(spacecraft)
[winner: 22434789 Soyuz_(rocket), pem_pos=1,
final_scores=-0.40769234, final_local_scores=-0.38696367,
pem=-1.0026959, similarity=-2.0052385|
gt: 178182 Soyuz_(spacecraft), pem_pos=4,
final_scores=-0.55652297, final_local_scores=-0.6017609,
pem=-1.8110224, similarity=-4.019411]
[winner: 6817421 Space_Shuttle_orbiter, pem_pos=4,
final_scores=-1.3522687, final_local_scores=-1.461082,
pem=-1.609438, similarity=-30.759514|

RECALL MISS: 178182 Soyuz_(spacecraft)]
10: 28237 Space_Shuttle_Columbia
[winner: 165108 Columbia_Records, pem_pos=0,
final_scores=-0.13495152, final_local_scores=-0.11202106,
pem=-1.516849, similarity=8.995642]
gt: 28237 Space_Shuttle_Columbia, pem_pos=1,
final_scores=-0.18872379, final_local_scores=-0.24972138,
pem=-2.4140012, similarity=9.765244]
11: 28189 Space_Shuttle
[winner: 94289 Sally_Ride, pem_pos=0,
final_scores=-0.7249364, final_local_scores=-0.81734294,
pem=0.0, similarity=-20.30143]
RECALL MISS: 28189 Space_Shuttle]
[winner: 1939119 Shuttle_(weaving), pem_pos=2,
final_scores=-1.0452148, final_local_scores=-0.9928652,
pem=-1.6171007, similarity=-16.74532]
gt: 28189 Space_Shuttle, pem_pos=0,
final_scores=-1.1392705, final_local_scores=-1.303808,
pem=-1.1889924, similarity=-28.352589]
12: 28235 Space_Shuttle_Challenger
[winner: 28235 Space_Shuttle_Challenger, pem_pos=0,
final_scores=-0.01742424, final_local_scores=-0.05530455,
pem=-0.62293684, similarity=5.826942]
gt: 28235 Space_Shuttle_Challenger, pem_pos=0,
final_scores=-0.01742424, final_local_scores=-0.05530455,
pem=-0.62293684, similarity=5.826942]
13: 26301 Rocket
[winner: 16984205 Armstrong_Siddeley_Screamer, pem_pos=6,
final_scores=-0.93251, final_local_scores=-0.87161714,
pem=-2.4731202, similarity=-8.4718685]
gt: 26301 Rocket, pem_pos=0, final_scores=-1.0395842,
final_local_scores=-1.1436298, pem=-0.15808994,
similarity=-29.178764]
14: 81326 Mir
[winner: 81326 Mir, pem_pos=0, final_scores=-0.034261823,
final_local_scores=-0.055644497, pem=-0.10771498,
similarity=3.014941]
gt: 81326 Mir, pem_pos=0, final_scores=-0.034261823,
final_local_scores=-0.055644497, pem=-0.10771498,
similarity=3.014941]
15: 28239 Space_Shuttle_Atlantis
[winner: 2178 Atlantis, pem_pos=0,
final_scores=-0.023317397, final_local_scores=0.0063758492,
pem=-0.6148956, similarity=7.6238785]
gt: 28239 Space_Shuttle_Atlantis, pem_pos=1,
final_scores=-0.12450275, final_local_scores=-0.16933334,
pem=-1.3361732, similarity=6.302786]
16: 28238 Space_Shuttle_Discovery
[winner: 28238 Space_Shuttle_Discovery, pem_pos=0,
final_scores=-0.07511808, final_local_scores=-0.11113949,
pem=-0.8705025, similarity=5.5070167]
gt: 28238 Space_Shuttle_Discovery, pem_pos=0,

```

final_scores=-0.07511808, final_local_scores=-0.11113949,
pem=-0.8705025, similarity=5.5070167 ]
1: 63727 Alan_Shepard
    [pem_pos=0, final_scores=0.3340035,
final_local_scores=0.30698374, pem=0.0, similarity=13.25072]
2: 58702 John_Glenn
    [pem_pos=0, final_scores=0.39824206,
final_local_scores=0.3635785, pem=-0.040958993,
similarity=15.162362]
3: 36592 Gus_Grissom
    [pem_pos=0, final_scores=0.2243542,
final_local_scores=0.2122891, pem=0.0, similarity=10.424843]
4: 21247 Neil_Armstrong
    [pem_pos=0, final_scores=0.33505106,
final_local_scores=0.32141158, pem=0.0, similarity=13.681275]
5: 65777 Buzz_Aldrin
    [pem_pos=0, final_scores=0.35521442,
final_local_scores=0.33964467, pem=0.0, similarity=14.225388]
6: 94289 Sally_Ride
    [pem_pos=0, final_scores=0.32806307,
final_local_scores=0.34211895, pem=0.0, similarity=14.299225]
7: 40203 Hubble_Space_Telescope
    [pem_pos=0, final_scores=0.19682293,
final_local_scores=0.19955707, pem=0.0, similarity=10.044894]
8: 514369 Norman_Thagard
    [pem_pos=0, final_scores=0.34444195,
final_local_scores=0.37172037, pem=0.0, similarity=15.18259]
9: 501402 Shannon_Lucid
    [pem_pos=0, final_scores=0.38031393,
final_local_scores=0.422629, pem=0.0, similarity=16.701805]
10: 380558 Story_Musgrave
    [pem_pos=0, final_scores=0.2689783,
final_local_scores=0.2853353, pem=0.0, similarity=12.604688]
1: 63727 Alan_Shepard
    [pem_pos=0, final_scores=0.12577917,
final_local_scores=0.07953031, pem=0.0, similarity=6.463056]
2: 58702 John_Glenn
    [pem_pos=0, final_scores=0.2162065,
final_local_scores=0.1647323, pem=-0.040958993,
similarity=9.2283945]
3: 3434750 United_States
    [pem_pos=1, final_scores=0.45773804,
final_local_scores=0.46414566, pem=-0.12994486,
similarity=18.647406]
4: 3434750 United_States
    [pem_pos=2, final_scores=0.40879327,
final_local_scores=0.40772027, pem=-0.12994486,
similarity=16.96356]
5: 3434750 United_States
    [pem_pos=1, final_scores=0.5183818,
final_local_scores=0.5303897, pem=-0.12994486,
similarity=20.624262]

```

6: 40203 Hubble_Space_Telescope
 [pem_pos=0, final_scores=0.092016056,
 final_local_scores=0.085071474, pem=-0.20751034,
 similarity=7.756889]
 7: 3434750 United_States
 [pem_pos=2, final_scores=0.36207986,
 final_local_scores=0.35965362, pem=-0.12994486,
 similarity=15.529154]
 8: 25391 Russia
 [pem_pos=1, final_scores=0.52130026,
 final_local_scores=0.5932214, pem=-1.0515022,
 similarity=27.510855]
 9: 3434750 United_States
 [pem_pos=2, final_scores=0.4007082,
 final_local_scores=0.4018492, pem=-0.12994486,
 similarity=16.788357]
 10: 81326 Mir
 [pem_pos=0, final_scores=0.09634994,
 final_local_scores=0.096783444, pem=-0.10771498,
 similarity=7.563695]
 11: 9228 Earth
 [pem_pos=0, final_scores=0.073939726,
 final_local_scores=0.08285655, pem=-0.07783105,
 similarity=6.985575]
 12: 81326 Mir
 [pem_pos=0, final_scores=0.08700113,
 final_local_scores=0.08657129, pem=-0.10771498,
 similarity=7.2589436]
 13: 3434750 United_States
 [pem_pos=0, final_scores=0.5709831,
 final_local_scores=0.5878487, pem=-0.14075686,
 similarity=22.397749]
 14: 58702 John_Glenn
 [pem_pos=0, final_scores=0.3667482,
 final_local_scores=0.32917622, pem=-0.040958993,
 similarity=14.135729]

Sample: b'APW19980624_0436_htm'

[1 Iraq 1] , [2 Iran 2] sign pact to resolve [3 POW 3] issue
 [1 1 Iraq 1] and [2 2 Iran 2] have signed a deal that [4 Iraq
 4] says will resolve their longstanding dispute over 1 prison
 ers of war . The agreement was reached following a week of
 talks in [3 3 Baghdad 3] between senior [5 Iraqi 5] officials
 and a visiting [6 Iranian 6] 2 delegation headed by Abdullah
 al-Najafi , head of [7 Iran 7] 's POW commission , the
 official [4 4 Iraqi News Agency 4] reported Wednesday . But
 the agency gave no further details and [8 al-Najafi 8] , who
 left [9 Baghdad 9] on Wednesday , declined to comment . Al-
 Najafi arrived in [10 Baghdad 10] last week with a list of
 thousands of [11 Iranian 11] POWs 3 Tehran claims are still
 held by [12 Iraq 12] . [13 Iraq 13] has long maintained that
 it has released all [14 Iranian 14] prisoners and was

reportedly angered by the new list . In April , [15 Iran 15] released 5,584 [16 Iraqi 16] POWs and [17 Iraq 17] freed three [18 POWs 18] and 316 [19 Iranians 19] it classified as civil law 4 detainees . It was the largest 5 repatriation of [20 Iranian 20] and [21 Iraqi 21] POWs since 1990 , said the [5 5 International Committee of the Red Cross 5] , which supervised the swap . The [22 Red Cross 22] says it knows of several thousand [23 Iraqis 23] in [24 Iran 24] , but it is not aware of more [25 Iranian 25] POWs in [26 Iraq 26] . The [6 6 Iran-Iraq 6] war , in which more than more than 1 million people were killed or wounded , ended in 1988 . The issue of [27 POWs 27] has long stood in the way of better ties . lb-as

1: 7515928 Iraq

2: 14653 Iran

3: 4492 Baghdad

4: 8550616 National_Iraqi_News_Agency

5: 5023832 International_Committee_of_the_Red_Cross

6: 14889 Iran-Iraq_War

1: 25008 Prisoner_of_war

[winner: 32611 Vietnam_War, pem_pos=0,
final_scores=-1.1671681, final_local_scores=-1.2713666,
pem=-0.12429769, similarity=-33.17445]

RECALL MISS: 25008 Prisoner_of_war]

[winner: 1600241 Prisoners'_rights, pem_pos=3,
final_scores=-1.3915089, final_local_scores=-1.4181893,
pem=-1.3818744, similarity=-30.717037]

gt: 25008 Prisoner_of_war, pem_pos=1,
final_scores=-1.519561, final_local_scores=-1.6725614,
pem=-0.71158856, similarity=-41.953125]

2: 218837 Delegation

[winner: 948582

United_States_congressional_delegations_from_Kansas,
pem_pos=6, final_scores=-1.2437664,
final_local_scores=-1.1663315, pem=-1.7882383,
similarity=-20.991232]

gt: 218837 Delegation, pem_pos=0,
final_scores=-1.7044156, final_local_scores=-1.7853923,
pem=-0.6624647, similarity=-45.587364]

3: 57654 Tehran

[winner: 14653 Iran, pem_pos=5, final_scores=-0.52332854,
final_local_scores=-0.6334978, pem=-2.6998966,
similarity=-0.13266665]

gt: 57654 Tehran, pem_pos=0, final_scores=-0.61630243,
final_local_scores=-0.7216138, pem=0.0,
similarity=-17.444683]

4: 5884870 Detention_(imprisonment)

[winner: 4406774 Guantanamo_Bay_detention_camp,
pem_pos=1, final_scores=-1.3753126,
final_local_scores=-1.4794004, pem=-0.54989314,
similarity=-37.06814]

gt: 5884870 Detention_(imprisonment), pem_pos=0,
final_scores=-1.3898687, final_local_scores=-1.4795527,

```

pem=0.0, similarity=-40.063095 ]
5: 1481067 Repatriation
    [winner: 1481067 Repatriation, pem_pos=0,
final_scores=-1.4790151, final_local_scores=-1.5984228,
pem=-0.23168783, similarity=-42.350456]
    gt: 1481067 Repatriation, pem_pos=0,
final_scores=-1.4790151, final_local_scores=-1.5984228,
pem=-0.23168783, similarity=-42.350456 ]
1: 7515928 Iraq
    [pem_pos=0, final_scores=0.5479745,
final_local_scores=0.49675357, pem=-0.25379938,
similarity=20.294031]
2: 14653 Iran
    [pem_pos=0, final_scores=0.6408135,
final_local_scores=0.642578, pem=-0.13856095,
similarity=24.01904]
3: 4492 Baghdad
    [pem_pos=0, final_scores=0.4076944,
final_local_scores=0.36577326, pem=0.0, similarity=15.005116]
4: 8550616 National_Iraqi_News_Agency
    [pem_pos=0, final_scores=0.35313702,
final_local_scores=0.53869796, pem=0.0, similarity=20.165535]
5: 5023832 International_Committee_of_the_Red_Cross
    [pem_pos=0, final_scores=0.4210469,
final_local_scores=0.45536685, pem=0.0, similarity=17.678768]
6: 14889 Iran-Iraq_War
    [pem_pos=0, final_scores=0.2555127,
final_local_scores=0.23163164, pem=0.0, similarity=11.002063]
1: 7515928 Iraq
    [pem_pos=0, final_scores=0.6051474,
final_local_scores=0.5592063, pem=-0.25379938,
similarity=22.157745]
2: 14653 Iran
    [pem_pos=0, final_scores=0.5407794,
final_local_scores=0.53330594, pem=-0.13856095,
similarity=20.758142]
3: 25008 Prisoner_of_war
    [pem_pos=0, final_scores=0.1591572,
final_local_scores=0.16833219, pem=0.0, similarity=9.113081]
4: 7515928 Iraq
    [pem_pos=0, final_scores=0.5456266,
final_local_scores=0.4941888, pem=-0.25379938,
similarity=20.217493]
5: 7515928 Iraq
    [pem_pos=0, final_scores=0.56856555,
final_local_scores=0.51924616, pem=-0.48469657,
similarity=22.22091]
6: 14653 Iran
    [pem_pos=0, final_scores=0.5475888,
final_local_scores=0.5407442, pem=-0.57066226,
similarity=23.329948]
7: 14653 Iran

```


[pem_pos=0, final_scores=0.66292006,
final_local_scores=0.666726, pem=-0.13856095,
similarity=24.739662]
8: 5537031 Usama_al-Nujayfi
[pem_pos=1, final_scores=0.08067392,
final_local_scores=0.17395759, pem=-0.6931472,
similarity=13.050396]
9: 4492 Baghdad
[pem_pos=0, final_scores=0.32247692,
final_local_scores=0.27268612, pem=0.0, similarity=12.227211]
10: 4492 Baghdad
[pem_pos=0, final_scores=0.4165221,
final_local_scores=0.37541616, pem=0.0, similarity=15.29288]
11: 14653 Iran
[pem_pos=0, final_scores=0.2823382,
final_local_scores=0.25099814, pem=-0.57066226,
similarity=14.683348]
12: 7515928 Iraq
[pem_pos=0, final_scores=0.61893195,
final_local_scores=0.5742638, pem=-0.25379938,
similarity=22.607092]
13: 7515928 Iraq
[pem_pos=0, final_scores=0.55813175,
final_local_scores=0.50784886, pem=-0.25379938,
similarity=20.625137]
14: 14653 Iran
[pem_pos=0, final_scores=0.5025719,
final_local_scores=0.49157006, pem=-0.57066226,
similarity=21.862494]
15: 14653 Iran
[pem_pos=0, final_scores=0.60428005,
final_local_scores=0.6026708, pem=-0.13856095,
similarity=22.828127]
16: 7515928 Iraq
[pem_pos=0, final_scores=0.37344915,
final_local_scores=0.3061111, pem=-0.48469657,
similarity=15.860534]
17: 7515928 Iraq
[pem_pos=0, final_scores=0.57107776,
final_local_scores=0.5219904, pem=-0.25379938,
similarity=21.04715]
18: 25008 Prisoner_of_war
[pem_pos=0, final_scores=0.065981254,
final_local_scores=0.066551626, pem=0.0, similarity=6.075746]
19: 14653 Iran
[pem_pos=1, final_scores=0.25243336,
final_local_scores=0.21833163, pem=-1.0731714,
similarity=16.441236]
20: 14653 Iran
[pem_pos=0, final_scores=0.40370005,
final_local_scores=0.38356757, pem=-0.57066226,
similarity=18.639484]

21: 7515928 Iraq
 [pem_pos=0, final_scores=0.35593444,
 final_local_scores=0.28697896, pem=-0.48469657,
 similarity=15.289593]
 22: 5023832 International_Committee_of_the_Red_Cross
 [pem_pos=1, final_scores=0.27035475,
 final_local_scores=0.29619217, pem=-1.9555908,
 similarity=23.563484]
 23: 7515928 Iraq
 [pem_pos=1, final_scores=0.3436579,
 final_local_scores=0.2735687, pem=-0.8129937,
 similarity=16.674734]
 24: 14653 Iran
 [pem_pos=0, final_scores=0.55033404,
 final_local_scores=0.54374295, pem=-0.13856095,
 similarity=21.069603]
 25: 14653 Iran
 [pem_pos=0, final_scores=0.3031081,
 final_local_scores=0.2736861, pem=-0.57066226,
 similarity=15.360401]
 26: 7515928 Iraq
 [pem_pos=0, final_scores=0.5971239,
 final_local_scores=0.5504418, pem=-0.25379938,
 similarity=21.896194]
 27: 25008 Prisoner_of_war
 [pem_pos=0, final_scores=0.15556099,
 final_local_scores=0.16440389, pem=0.0, similarity=8.995852]

Sample: b'APW19981106_0920_htm'

[1 Norwegian 1] company moves ahead with clean power plant
 The [1 1 Norsk Hydro ASA 1] concern announced plans Friday to
 build a 1 hydrogen - burning 2 power plant that it says will
 produce 3 electricity with virtually no climate-damaging 4 em
 issions . The 51-percent 5 state-owned company notified the 6
 Norwegian Water Resources and Energy Directorate of plans to
 build the plant , to be called HydroKraft , pending approval
 by government agencies . The company said if it receives the
 proper approvals in 1999 , construction of the plant would be
 completed during 2002 and production startup would occur in
 early 2003 . Earlier this year , the company announced a new
 technology that would allow it to use clean-burning hydrogen
 to generate electricity . The hydrogen for the plant would be
 extracted from 7 natural gas . [2 Norsk Hydro 2] , a major 8 p
 etroleum producer , now says studies have shown the proposal
 to be economically viable . It also said the plant 's 9 carbon
 dioxide emissions would be 90 percent lower than plants using
 natural gas . [2 2 Norway 2] is the world 's second-largest
 oil exporter , and also produces vast amounts of natural gas
 from its offshore fields . Virtually all of the country 's
 electricity now comes from 10 hydroelectric plants . Under
 the company 's plans , natural gas , air and water would be
 mixed to produce hydrogen , 11 nitrogen and carbon dioxide . [

3 Norsk Hydro 3] wants to build the plant at Karmøy , an island about 500 kilometers -LRB- 300 miles -RRB- west of [3 3 Oslo 3] .

1: 337551 Norsk_Hydro

2: 21241 Norway

3: 22309 Oslo

1: 13255 Hydrogen

[winner: 18504615 Hydrogen_odorant, pem_pos=3, final_scores=-0.9000157, final_local_scores=-0.79354435, pem=-1.3848393, similarity=-12.060261|

gt: 13255 Hydrogen, pem_pos=0, final_scores=-1.081524, final_local_scores=-1.0712487, pem=-0.06824774, similarity=-27.507341]

2: 212141 Power_station

[winner: 9695679 Central_Islip_State_Hospital_Powerplant, pem_pos=5, final_scores=-1.0282379, final_local_scores=-0.95974255, pem=-1.6081792, similarity=-15.805389|

gt: 212141 Power_station, pem_pos=0, final_scores=-1.2756655, final_local_scores=-1.3466566, pem=-0.08021067, similarity=-35.661007]

3: 9550 Electricity

[winner: 225268 Electric_eel, pem_pos=0, final_scores=-0.99491477, final_local_scores=-0.8891574, pem=0.0, similarity=-22.444517|

RECALL MISS: 9550 Electricity]

4: 10934212 Air_pollution

[winner: 840147 Exhaust_gas, pem_pos=1, final_scores=-1.0735004, final_local_scores=-1.0455482, pem=-1.1069607, similarity=-21.09171|

gt: 10934212 Air_pollution, pem_pos=3, final_scores=-1.3428698, final_local_scores=-1.3765695, pem=-1.405964, similarity=-29.344013]

5: 204768 Government-owned_corporation

[winner: 204768 Government-owned_corporation, pem_pos=0, final_scores=-1.0998209, final_local_scores=-1.1734691, pem=0.0, similarity=-30.928944|

gt: 204768 Government-owned_corporation, pem_pos=0, final_scores=-1.0998209, final_local_scores=-1.1734691, pem=0.0, similarity=-30.928944]

6: 4596009 Norwegian_Water_Resources_and_Energy_Directorate

[winner: 4596009

Norwegian_Water_Resources_and_Energy_Directorate, pem_pos=0, final_scores=-0.11440356, final_local_scores=-0.10755117, pem=0.0, similarity=0.8801713|

gt: 4596009

Norwegian_Water_Resources_and_Energy_Directorate, pem_pos=0, final_scores=-0.11440356, final_local_scores=-0.10755117, pem=0.0, similarity=0.8801713]

7: 22131 Natural_gas

[winner: 18993869 Gas, pem_pos=0, final_scores=-1.0891609, final_local_scores=-1.1232731,

pem=-0.3543779, similarity=-27.503834|
gt: 22131 Natural_gas, pem_pos=1,
final_scores=-1.3427255, final_local_scores=-1.4416275,
pem=-1.4513136, similarity=-31.038862]
[winner: 6055139 NATURAL, pem_pos=15,
final_scores=-1.2070776, final_local_scores=-1.1209402,
pem=-3.6564257, similarity=-9.477165|
gt: 22131 Natural_gas, pem_pos=5,
final_scores=-1.5857509, final_local_scores=-1.7070959,
pem=-2.3708131, similarity=-33.960587]
8: 23195 Petroleum
[winner: 271860 Petroleum_jelly, pem_pos=0,
final_scores=-1.0373188, final_local_scores=-1.0001674,
pem=0.0, similarity=-25.757275|
gt: 23195 Petroleum, pem_pos=1, final_scores=-1.3131714,
final_local_scores=-1.4072411, pem=-0.015366355,
similarity=-37.821606]
9: 5906 Carbon_dioxide
[winner: 2951653 Supercritical_carbon_dioxide, pem_pos=2,
final_scores=-0.93703276, final_local_scores=-0.8566665,
pem=-0.6931472, similarity=-17.705482|
gt: 5906 Carbon_dioxide, pem_pos=0,
final_scores=-1.194179, final_local_scores=-1.2017813,
pem=-0.02332073, similarity=-31.647017]
10: 381399 Hydroelectricity
[winner: 14073 Hydropower, pem_pos=0,
final_scores=-1.3181171, final_local_scores=-1.3860011,
pem=0.0, similarity=-37.271328|
gt: 381399 Hydroelectricity, pem_pos=1,
final_scores=-1.4637889, final_local_scores=-1.5438519,
pem=-0.021431543, similarity=-41.86536]
[winner: 3907148 Generation_III_reactor, pem_pos=4,
final_scores=-1.3201727, final_local_scores=-1.3062518,
pem=-1.3862944, similarity=-27.352558|
RECALL MISS: 381399 Hydroelectricity]
11: 21175 Nitrogen
[winner: 21175 Nitrogen, pem_pos=1,
final_scores=-1.06104, final_local_scores=-1.0341331,
pem=-0.016324058, similarity=-26.682108|
gt: 21175 Nitrogen, pem_pos=1, final_scores=-1.06104,
final_local_scores=-1.0341331, pem=-0.016324058,
similarity=-26.682108]
1: 337551 Norsk_Hydro
[pem_pos=0, final_scores=0.3254941,
final_local_scores=0.28119546, pem=0.0, similarity=12.481147]
2: 21241 Norway
[pem_pos=0, final_scores=0.64174277,
final_local_scores=0.65468854, pem=-0.22668563,
similarity=24.859676]
3: 22309 Oslo
[pem_pos=0, final_scores=0.5441031,
final_local_scores=0.5625335, pem=-0.035297178,

similarity=21.068785]

1: 21241 Norway

[pem_pos=0, final_scores=0.53736186,
final_local_scores=0.5406682, pem=-0.605967,
similarity=23.519672]

2: 337551 Norsk_Hydro

[pem_pos=0, final_scores=0.36563438,
final_local_scores=0.31959197, pem=0.0, similarity=13.626975]

3: 337551 Norsk_Hydro

[pem_pos=0, final_scores=0.3272112,
final_local_scores=0.27762043, pem=0.0, similarity=12.37446]

Sample: b'APW19981022_0630_htm'

Interpol needs to revamp **1 strategy** to keep abreast of crime
2 Interpol must modernize if it wants to effectively fight
increasingly sophisticated **3 crime syndicates** in the next
century , the police organization 's president said
Thursday . The criminal environment has changed at a `` much
greater pace '' than Interpol , Toshinori Kanemoto told the
opening session of the 67th general assembly of the **[1**
International Criminal Police Organization **1]** . A failure to
modernize Interpol could mean the organization `` risks being
displaced against a backdrop of international **4 organized cri**
me , '' he said . Kanemoto said a new **[2 Interpol 2]** strategy
will be discussed at the six-day assembly and , if adopted ,
would be put into effect next January . The former **[3**
Japanese **3]** police official said the assembly also will
discuss **5 terrorism** , which he called one of the `` new forms
of criminality , '' as well as the **6 illegal drug trade** and **7**
money laundering . With crime becoming increasingly
sophisticated , **[4 Kanemoto 4]** said , it has become harder
for police officials to gather **8 evidence** for prosecution .
He said these new challenges require increased cooperation ,
particularly through Interpol . The police officials are
expected to discuss a **9 strategic development plan** aimed at
making Interpol more proactive by re-examining the needs of
the group 's 177-member states , **[5 Kanemoto 5]** said .

Details of , graf 8 pvs

1: 29607 Strategy

[winner: 164865 Strategy_pattern, pem_pos=2,
final_scores=-1.3125044, final_local_scores=-1.2740104,
pem=-1.2231199, similarity=-27.27778]

gt: 29607 Strategy, pem_pos=0, final_scores=-1.4981375,
final_local_scores=-1.5529096, pem=-0.56880796,
similarity=-39.15894]

2: 14338129 Interpol

[winner: 14338129 Interpol, pem_pos=0,
final_scores=-0.07313809, final_local_scores=-0.07058177,
pem=-0.41894695, similarity=4.2617106]

gt: 14338129 Interpol, pem_pos=0,
final_scores=-0.07313809, final_local_scores=-0.07058177,
pem=-0.41894695, similarity=4.2617106]

3: 22625 Organized_crime
[winner: 3304973 Organised_crime_in_India, pem_pos=2,
final_scores=-1.1813883, final_local_scores=-1.151432,
pem=-0.6618946, similarity=-26.671831|
gt: 22625 Organized_crime, pem_pos=1,
final_scores=-1.4105563, final_local_scores=-1.492285,
pem=-0.5458225, similarity=-37.474785]
4: 22625 Organized_crime
[winner: 5841680 Transnational_organized_crime,
pem_pos=0, final_scores=-0.8081112,
final_local_scores=-0.80308694, pem=-0.105360545,
similarity=-19.303036|
gt: 22625 Organized_crime, pem_pos=1,
final_scores=-1.4828212, final_local_scores=-1.5712235,
pem=-2.3025851, similarity=-30.276915]
5: 30636 Terrorism
[winner: 30636 Terrorism, pem_pos=0,
final_scores=-1.0875832, final_local_scores=-1.1381223,
pem=-0.18121491, similarity=-28.888655|
gt: 30636 Terrorism, pem_pos=0, final_scores=-1.0875832,
final_local_scores=-1.1381223, pem=-0.18121491,
similarity=-28.888655]
6: 218449 Illegal_drug_trade
[winner: 803469 ATC_code_N, pem_pos=9,
final_scores=-0.86936057, final_local_scores=-0.69544935,
pem=-2.9199536, similarity=-0.7847218|
RECALL MISS: 218449 Illegal_drug_trade]
[winner: 465780 Illegal_logging, pem_pos=1,
final_scores=-1.4396075, final_local_scores=-1.4456267,
pem=-1.7986251, similarity=-29.269472|
gt: 218449 Illegal_drug_trade, pem_pos=16,
final_scores=-2.0756478, final_local_scores=-2.217075,
pem=-3.9149249, similarity=-40.782265]
[winner: 1090151 Border_trade, pem_pos=5,
final_scores=-1.5081992, final_local_scores=-1.5071254,
pem=-2.0768182, similarity=-29.591852|
RECALL MISS: 218449 Illegal_drug_trade]
7: 19390 Money_laundering
[winner: 19390 Money_laundering, pem_pos=0,
final_scores=-1.0908636, final_local_scores=-1.1297543,
pem=-0.026342224, similarity=-29.481155|
gt: 19390 Money_laundering, pem_pos=0,
final_scores=-1.0908636, final_local_scores=-1.1297543,
pem=-0.026342224, similarity=-29.481155]
8: 20550772 Evidence
[winner: 16470771 Entomological_evidence_collection,
pem_pos=0, final_scores=-1.218765,
final_local_scores=-1.1688993, pem=0.0, similarity=-30.79257|
RECALL MISS: 20550772 Evidence]
9: 227101 Strategic_planning
[winner: 3873661 Plans_in_Mexican_history, pem_pos=4,
final_scores=-1.3152225, final_local_scores=-1.2484608,

```

pem=-1.7760347, similarity=-23.508495|
  RECALL MISS: 227101 Strategic_planning]
  [winner: 28118 Strategic_Air_Command, pem_pos=4,
final_scores=-1.3211231, final_local_scores=-1.3534901,
pem=-1.6081679, similarity=-27.555666|
  gt: 227101 Strategic_planning, pem_pos=10,
final_scores=-2.2059023, final_local_scores=-2.3040895,
pem=-5.209486, similarity=-36.33892 ]
  [winner: 555622 Musical_development, pem_pos=2,
final_scores=-1.5658563, final_local_scores=-1.5145183,
pem=-2.3828495, similarity=-28.148228|
  RECALL MISS: 227101 Strategic_planning]
1: 14338129 Interpol
  [pem_pos=0, final_scores=0.045860156,
final_local_scores=0.09842053, pem=0.0, similarity=7.0267777]
2: 14338129 Interpol
  [pem_pos=0, final_scores=0.06303696,
final_local_scores=0.117183596, pem=-0.41894695,
similarity=9.865004]
3: 15573 Japan
  [pem_pos=0, final_scores=0.7502001,
final_local_scores=0.85017043, pem=-0.99076414,
similarity=34.848423]
4: 22356629 Tomoaki_Kanemoto
  [pem_pos=0, final_scores=0.051939234,
final_local_scores=0.06333175, pem=-0.37806612,
similarity=8.035641]
5: 22356629 Tomoaki_Kanemoto
  [pem_pos=0, final_scores=0.04681106,
final_local_scores=0.045098126, pem=-0.37806612,
similarity=7.4915133]

```

Sample: b'APW19990526_0131_htm'

Alaska Hiker Killed by 1 Bear [1 1 ANCHORAGE 1] , [2 2 Alaska 2] -LRB- AP -RRB- -- A man on a day hike in a rural area was killed when a bear bit his head , [3 3 Alaska State Troopers 3] said Wednesday . Kenneth Cates , 53 , was found Wednesday morning on a horse trail in a heavily wooded area of the [4 4 Kenai National Wildlife Refuge 4] near his hometown of [5 5 So ldotna 5] . The attack happened about six miles off a busy road . The area is about 60 miles southwest of [1 Anchorage 1] . The fatal mauling is thought to have happened Tuesday . His only visible wound was a single bite to the head , said Bruce Bartley , a spokesman for the state Division of Wildlife Conservation . State and federal wildlife officials were looking for the bear involved in the attack . If the bear is judged to pose a continuing threat to humans it may be killed . Troopers said Cates ' .280 - caliber 2 rifle and two spent shell casings were found near the body , along with traces of blood that suggest that Cates may have wounded the animal . `` We do n't know if one was a warning shot and one real , or if there were two real shots , '' Bartley said .

Cates ' death is the first fatal bear mauling in [2 Alaska 2] since Feb. 8 , 1998 , when Audelio Luis Cortes , 40 , was killed while working on a 3 seismic crew in the [6 6 Swanson River 6] 4 oil field near [7 7 Kenai 7] . Cortes also died from a single head bite .

1: 7004698 Anchorage,_Alaska

2: 624 Alaska

3: 2645175 Alaska_State_Troopers

4: 8846958 Kenai_National_Wildlife_Refuge

5: 105492 Soldotna,_Alaska

6: 988425 Swanson_River

7: 105477 Kenai,_Alaska

1: 4400 Bear

[winner: 4400 Bear, pem_pos=0, final_scores=-0.60064423, final_local_scores=-0.61477053, pem=-0.74816996, similarity=-10.187602]

gt: 4400 Bear, pem_pos=0, final_scores=-0.60064423, final_local_scores=-0.61477053, pem=-0.74816996, similarity=-10.187602]

2: 25456 Rifle

[winner: 11378031 Chropi_rifle, pem_pos=6, final_scores=-1.053032, final_local_scores=-0.93547624, pem=-2.0792258, similarity=-12.51961]

gt: 25456 Rifle, pem_pos=0, final_scores=-1.082609, final_local_scores=-1.1162425, pem=-0.00934778, similarity=-29.170357]

3: 28935 Seismology

[winner: 24762328 Seismic_metamaterials, pem_pos=0, final_scores=-0.74060506, final_local_scores=-0.6370074, pem=0.0, similarity=-14.919859]

gt: 28935 Seismology, pem_pos=1, final_scores=-1.1816809, final_local_scores=-1.2244782, pem=-0.10388304, similarity=-31.886225]

4: 219931 Oil_field

[winner: 1075931 Brae_oilfield, pem_pos=2, final_scores=-1.1011739, final_local_scores=-1.0402904, pem=-1.7895356, similarity=-17.22285]

gt: 219931 Oil_field, pem_pos=0, final_scores=-1.5286034, final_local_scores=-1.605753, pem=-0.02024058, similarity=-43.71909]

1: 7004698 Anchorage,_Alaska

[pem_pos=0, final_scores=0.51331073, final_local_scores=0.47609532, pem=-0.046072915, similarity=18.5479]

2: 624 Alaska

[pem_pos=0, final_scores=0.48023278, final_local_scores=0.4450845, pem=-0.18647984, similarity=18.386028]

3: 2645175 Alaska_State_Troopers

[pem_pos=0, final_scores=0.17065696, final_local_scores=0.19558787, pem=0.0, similarity=9.926445]

4: 8846958 Kenai_National_Wildlife_Refuge

[pem_pos=0, final_scores=0.32367706,
 final_local_scores=0.3880484, pem=0.0, similarity=15.669851]
 5: 105492 Soldotna, Alaska
 [pem_pos=0, final_scores=0.3008023,
 final_local_scores=0.2931453, pem=0.0, similarity=12.837753]
 6: 988425 Swanson River
 [pem_pos=0, final_scores=0.18291746,
 final_local_scores=0.25632903, pem=0.0, similarity=11.739082]
 7: 105477 Kenai, Alaska
 [pem_pos=0, final_scores=0.34921223,
 final_local_scores=0.33177224, pem=-0.07768023,
 similarity=14.412896]
 1: 7004698 Anchorage, Alaska
 [pem_pos=0, final_scores=0.48735064,
 final_local_scores=0.44773787, pem=-0.046072915,
 similarity=17.701656]
 2: 624 Alaska
 [pem_pos=0, final_scores=0.60240656,
 final_local_scores=0.57854086, pem=-0.18647984,
 similarity=22.368631]

Sample: b'APW19980808_0196_htm'

[1 Saddam 1] says [2 U.S. 2] will reap nothing , urges [3
 Iran 3] to normalize ties President [1 1 Saddam Hussein 1]
 said Saturday the [2 2 United States 2] will reap nothing but
 a `` harvest ... full of thorns '' from its campaign against [3
 3 3 Iraq 3] . Addressing the nation on television to mark the
 anniversary of the end of the 1980-88 [4 4 Iran-Iraq 4] war ,
 [4 Saddam 4] also urged [5 Iran 5] to normalize relations
 with [6 Iraq 6] . [7 Saddam 7] did not refer to his decision
 Wednesday to freeze cooperation with [5 5 U.N. 5] arms
 inspectors , but he said the [8 United States 8] and its 1 al
 lies would be misled if they thought they could bring [9 Iraq
 9] to its knees . This is exactly what the [10 2 Iranian 10] g
 overnment under the late [6 6 Ayatollah Ruhollah Khomeini 6]
 thought it could do in the war , but it failed , [11 Saddam 11
] said . `` Those imposing aggression on [12 Iraq 12] ... and
 the unfair 3 embargo neglect this fact today , '' [13 Saddam 1
 3] said , referring to sweeping 4 sanctions the [14 United
 Nations 14] imposed after [15 Iraq 15] 's 1990 invasion of [7
 7 Kuwait 7] . `` They are making the same mistake . Their
 harvest will be full of thorns and the outcome will be a
 taste of bitterness . '' The president delivered his speech
 in a black suit . He had decreed that Saturday be celebrated
 as a day of victory over [16 Iran 16] , and an 5 artillery
 salute of 101 guns boomed over [8 8 Baghdad 8] in the early
 morning . National radio and television played 6 patriotic
 songs for the occasion . [17 Saddam 17] spoke , 7th graf pvs
 1: 29490 Saddam_Hussein
 2: 3434750 United_States
 3: 7515928 Iraq
 4: 14889 Iran-Iraq_War

5: 31769 United_Nations
 6: 26232 Ruhollah_Khomeini
 7: 7515890 Kuwait
 8: 4492 Baghdad
 1: 2198844 Allies_of_World_War_II
 [winner: 6574080 Allies_(Crosby,_Stills_&_Nash_album),
 pem_pos=6, final_scores=-1.2638705,
 final_local_scores=-1.2049029, pem=-2.3417265,
 similarity=-19.132317|
 gt: 2198844 Allies_of_World_War_II, pem_pos=0,
 final_scores=-1.4577403, final_local_scores=-1.6017275,
 pem=-0.47413486, similarity=-41.130608]
 2: 64702 Politics_of_Iran
 [winner: 14653 Iran, pem_pos=0, final_scores=0.49034053,
 final_local_scores=0.52518135, pem=-0.57066226,
 similarity=22.865522|
 RECALL MISS: 64702 Politics_of_Iran]
 [winner: 12229 Government, pem_pos=0,
 final_scores=-1.6407021, final_local_scores=-1.7829652,
 pem=-0.3899827, similarity=-46.99674|
 RECALL MISS: 64702 Politics_of_Iran]
 3: 61660 Embargo
 [winner: 2030237 United_States_embargoes, pem_pos=3,
 final_scores=-0.9259376, final_local_scores=-0.9147065,
 pem=-1.0898758, similarity=-17.28004|
 gt: 61660 Embargo, pem_pos=0, final_scores=-1.1180044,
 final_local_scores=-1.2197456, pem=-0.3612818,
 similarity=-30.345224]
 4: 39411680 International_sanctions
 [winner: 14469390
 United_Nations_Security_Council_Resolution_282, pem_pos=4,
 final_scores=-0.99772704, final_local_scores=-0.9886417,
 pem=-1.7893898, similarity=-15.682344|
 gt: 39411680 International_sanctions, pem_pos=0,
 final_scores=-1.0591583, final_local_scores=-1.1528707,
 pem=-0.629573, similarity=-26.890528]
 5: 2508 Artillery
 [winner: 2508 Artillery, pem_pos=0,
 final_scores=-1.2302483, final_local_scores=-1.3104265,
 pem=-0.04625687, similarity=-34.764477|
 gt: 2508 Artillery, pem_pos=0, final_scores=-1.2302483,
 final_local_scores=-1.3104265, pem=-0.04625687,
 similarity=-34.764477]
 6: 24552 Patriotism
 [winner: 26250614
 Patriot_(Spanish_American_Independence), pem_pos=1,
 final_scores=-1.1827857, final_local_scores=-1.1705852,
 pem=-0.6871344, similarity=-27.106142|
 gt: 24552 Patriotism, pem_pos=0,
 final_scores=-1.3413193, final_local_scores=-1.4382224,
 pem=-0.059400313, similarity=-38.506687]
 1: 29490 Saddam_Hussein

[pem_pos=0, final_scores=0.5256113,
final_local_scores=0.48940623, pem=0.0, similarity=18.69457]
2: 3434750 United_States
[pem_pos=0, final_scores=0.5027787,
final_local_scores=0.51911503, pem=-0.073457584,
similarity=19.980616]
3: 7515928 Iraq
[pem_pos=0, final_scores=0.61220676,
final_local_scores=0.56505257, pem=-0.25379938,
similarity=22.332209]
4: 14889 Iran-Iraq_War
[pem_pos=0, final_scores=0.30883646,
final_local_scores=0.27191538, pem=0.0, similarity=12.20421]
5: 31769 United_Nations
[pem_pos=0, final_scores=0.48703104,
final_local_scores=0.5012515, pem=0.0, similarity=19.048058]
6: 26232 Ruhollah_Khomeini
[pem_pos=0, final_scores=0.40480244,
final_local_scores=0.40735215, pem=0.0, similarity=16.245914]
7: 7515890 Kuwait
[pem_pos=0, final_scores=0.4908207,
final_local_scores=0.4600436, pem=-0.3372397,
similarity=19.652294]
8: 4492 Baghdad
[pem_pos=0, final_scores=0.32433933,
final_local_scores=0.2778191, pem=0.0, similarity=12.380389]
1: 29490 Saddam_Hussein
[pem_pos=0, final_scores=0.3747546,
final_local_scores=0.32461822, pem=0.0, similarity=13.776968]
2: 3434750 United_States
[pem_pos=0, final_scores=0.45742875,
final_local_scores=0.46957713, pem=-0.14075686,
similarity=18.86829]
3: 14653 Iran
[pem_pos=0, final_scores=0.5680033,
final_local_scores=0.61001617, pem=-0.13856095,
similarity=23.047329]
4: 29490 Saddam_Hussein
[pem_pos=0, final_scores=0.4235499,
final_local_scores=0.37791967, pem=0.0, similarity=15.367589]
5: 14653 Iran
[pem_pos=0, final_scores=0.6037272,
final_local_scores=0.64903915, pem=-0.13856095,
similarity=24.211851]
6: 7515928 Iraq
[pem_pos=0, final_scores=0.6686633,
final_local_scores=0.6267228, pem=-0.25379938,
similarity=24.172573]
7: 29490 Saddam_Hussein
[pem_pos=0, final_scores=0.35441393,
final_local_scores=0.30239916, pem=0.0, similarity=13.113907]
8: 3434750 United_States

[pem_pos=0, final_scores=0.48970133,
 final_local_scores=0.50483, pem=-0.073457584,
 similarity=19.554321]
 9: 7515928 Iraq
 [pem_pos=0, final_scores=0.53442496,
 final_local_scores=0.48008776, pem=-0.25379938,
 similarity=19.79669]
 10: 14653 Iran
 [pem_pos=0, final_scores=0.49034053,
 final_local_scores=0.52518135, pem=-0.57066226,
 similarity=22.865522]
 11: 29490 Saddam_Hussein
 [pem_pos=0, final_scores=0.3725304,
 final_local_scores=0.3221886, pem=0.0, similarity=13.704463]
 12: 7515928 Iraq
 [pem_pos=0, final_scores=0.56617016,
 final_local_scores=0.5147646, pem=-0.25379938,
 similarity=20.831516]
 13: 29490 Saddam_Hussein
 [pem_pos=0, final_scores=0.39008325,
 final_local_scores=0.34136245, pem=0.0, similarity=14.276649]
 14: 31769 United_Nations
 [pem_pos=0, final_scores=0.42723876,
 final_local_scores=0.43593746, pem=0.0, similarity=17.098957]
 15: 7515928 Iraq
 [pem_pos=0, final_scores=0.5872542,
 final_local_scores=0.5377957, pem=-0.25379938,
 similarity=21.51881]
 16: 14653 Iran
 [pem_pos=0, final_scores=0.54991245,
 final_local_scores=0.59025466, pem=-0.13856095,
 similarity=22.457605]
 17: 29490 Saddam_Hussein
 [pem_pos=0, final_scores=0.36485022,
 final_local_scores=0.31379917, pem=0.0, similarity=13.454106]

Sample: b'APW19980624_0607_htm'

[1 Iraq 1] , [2 Iran 2] sign pact to resolve [3 POW 3] issue
 [1 1 Iraq 1] and [2 2 Iran 2] have signed a deal that [4 Iraq
 4] says will resolve their longstanding dispute over 1 prison
 ers of war . The agreement was reached following a week of
 talks in [3 3 Baghdad 3] between senior [5 Iraqi 5] officials
 and a visiting [6 Iranian 6] delegation headed by Abdullah al-
 Najafi , head of [7 Iran 7] 's POW commission , the official [4
 4 Iraqi News Agency 4] reported Wednesday . But the agency
 gave no further details and [8 al-Najafi 8] , who left [9
 Baghdad 9] on Wednesday , declined to comment . Al-Najafi
 arrived in [10 Baghdad 10] last week with a list of thousands
 of [11 Iranian 11] POWs 2 Tehran claims are still held by [12
 Iraq 12] . [13 Iraq 13] has long maintained that it has
 released all [14 Iranian 14] prisoners and was reportedly
 angered by the new list . In April , [15 Iran 15] released

5,584 [16 Iraqi 16] POWs and [17 Iraq 17] freed three [18 POWs 18] and 316 [19 Iranians 19] it classified as civil law detainees . It was the largest 3 repatriation of [20 Iranian 20] and [21 Iraqi 21] POWs since 1990 , said the [5 5 International Committee of the Red Cross 5] , which supervised the swap . The [22 Red Cross 22] says it knows of several thousand [23 Iraqis 23] in [24 Iran 24] , but it is not aware of more [25 Iranian 25] POWs in [26 Iraq 26] . The [6 6 Iran-Iraq 6] war , in which more than more than 1 million people were killed or wounded , ended in 1988 . The issue of [27 POWs 27] has long stood in the way of better ties .

1: 7515928 Iraq

2: 14653 Iran

3: 4492 Baghdad

4: 8550616 National_Iraqi_News_Agency

5: 5023832 International_Committee_of_the_Red_Cross

6: 14889 Iran-Iraq_War

1: 25008 Prisoner_of_war

[winner: 32611 Vietnam_War, pem_pos=0,
final_scores=-1.167167, final_local_scores=-1.2713654,
pem=-0.12429769, similarity=-33.174416]

RECALL MISS: 25008 Prisoner_of_war]

[winner: 1600241 Prisoners'_rights, pem_pos=3,
final_scores=-1.3915089, final_local_scores=-1.4181893,
pem=-1.3818744, similarity=-30.717037]

gt: 25008 Prisoner_of_war, pem_pos=1,
final_scores=-1.5195608, final_local_scores=-1.6725612,
pem=-0.71158856, similarity=-41.95312]

2: 57654 Tehran

[winner: 14653 Iran, pem_pos=5, final_scores=-0.52322125,
final_local_scores=-0.63338053, pem=-2.6998966,
similarity=-0.12916781]

gt: 57654 Tehran, pem_pos=0, final_scores=-0.616182,
final_local_scores=-0.7214823, pem=0.0,
similarity=-17.440758]

3: 1481067 Repatriation

[winner: 1481067 Repatriation, pem_pos=0,
final_scores=-1.4790162, final_local_scores=-1.598424,
pem=-0.23168783, similarity=-42.350494]

gt: 1481067 Repatriation, pem_pos=0,
final_scores=-1.4790162, final_local_scores=-1.598424,
pem=-0.23168783, similarity=-42.350494]

1: 7515928 Iraq

[pem_pos=0, final_scores=0.5490739,
final_local_scores=0.49795443, pem=-0.25379938,
similarity=20.329866]

2: 14653 Iran

[pem_pos=0, final_scores=0.6418933,
final_local_scores=0.6437575, pem=-0.13856095,
similarity=24.054237]

3: 4492 Baghdad

[pem_pos=0, final_scores=0.40783757,

final_local_scores=0.36592966, pem=0.0, similarity=15.009785]
4: 8550616 National_Iraqi_News_Agency
[pem_pos=0, final_scores=0.35315627,
final_local_scores=0.538719, pem=0.0, similarity=20.166162]
5: 5023832 International_Committee_of_the_Red_Cross
[pem_pos=0, final_scores=0.4210624,
final_local_scores=0.45538378, pem=0.0, similarity=17.679274]
6: 14889 Iran-Iraq_War
[pem_pos=0, final_scores=0.2558341,
final_local_scores=0.23198271, pem=0.0, similarity=11.01254]
1: 7515928 Iraq
[pem_pos=0, final_scores=0.6058101,
final_local_scores=0.55993015, pem=-0.25379938,
similarity=22.179346]
2: 14653 Iran
[pem_pos=0, final_scores=0.54288083,
final_local_scores=0.53560144, pem=-0.13856095,
similarity=20.826645]
3: 25008 Prisoner_of_war
[pem_pos=0, final_scores=0.16012003,
final_local_scores=0.16938394, pem=0.0, similarity=9.144467]
4: 7515928 Iraq
[pem_pos=0, final_scores=0.5462217,
final_local_scores=0.49483883, pem=-0.25379938,
similarity=20.23689]
5: 7515928 Iraq
[pem_pos=0, final_scores=0.57294387,
final_local_scores=0.52402884, pem=-0.48469657,
similarity=22.363634]
6: 14653 Iran
[pem_pos=0, final_scores=0.55029374,
final_local_scores=0.54369885, pem=-0.57066226,
similarity=23.418121]
7: 14653 Iran
[pem_pos=0, final_scores=0.6629289,
final_local_scores=0.6667357, pem=-0.13856095,
similarity=24.739952]
8: 5537031 Usama_al-Nujayfi
[pem_pos=1, final_scores=0.0808572,
final_local_scores=0.1741578, pem=-0.6931472,
similarity=13.056371]
9: 4492 Baghdad
[pem_pos=0, final_scores=0.32256347,
final_local_scores=0.2727807, pem=0.0, similarity=12.230033]
10: 4492 Baghdad
[pem_pos=0, final_scores=0.4167269,
final_local_scores=0.37563986, pem=0.0, similarity=15.299555]
11: 14653 Iran
[pem_pos=0, final_scores=0.28256708,
final_local_scores=0.25124818, pem=-0.57066226,
similarity=14.69081]
12: 7515928 Iraq

[pem_pos=0, final_scores=0.6206134,
final_local_scores=0.5761005, pem=-0.25379938,
similarity=22.661903]
13: 7515928 Iraq
[pem_pos=0, final_scores=0.559202,
final_local_scores=0.5090179, pem=-0.25379938,
similarity=20.660023]
14: 14653 Iran
[pem_pos=0, final_scores=0.5032127,
final_local_scores=0.49227, pem=-0.57066226,
similarity=21.88338]
15: 14653 Iran
[pem_pos=0, final_scores=0.60527617,
final_local_scores=0.6037588, pem=-0.13856095,
similarity=22.860596]
16: 7515928 Iraq
[pem_pos=0, final_scores=0.37424058,
final_local_scores=0.3069756, pem=-0.48469657,
similarity=15.8863325]
17: 7515928 Iraq
[pem_pos=0, final_scores=0.571873,
final_local_scores=0.52285904, pem=-0.25379938,
similarity=21.07307]
18: 25008 Prisoner_of_war
[pem_pos=0, final_scores=0.066846296,
final_local_scores=0.06749655, pem=0.0, similarity=6.103945]
19: 14653 Iran
[pem_pos=1, final_scores=0.25251102,
final_local_scores=0.21841651, pem=-1.0731714,
similarity=16.44377]
20: 14653 Iran
[pem_pos=0, final_scores=0.4050042,
final_local_scores=0.38499212, pem=-0.57066226,
similarity=18.681995]
21: 7515928 Iraq
[pem_pos=0, final_scores=0.35593736,
final_local_scores=0.28698218, pem=-0.48469657,
similarity=15.289688]
22: 5023832 International_Committee_of_the_Red_Cross
[pem_pos=1, final_scores=0.27075887,
final_local_scores=0.2966336, pem=-1.9555908,
similarity=23.576658]
23: 7515928 Iraq
[pem_pos=1, final_scores=0.34405738,
final_local_scores=0.27400512, pem=-0.8129937,
similarity=16.68776]
24: 14653 Iran
[pem_pos=0, final_scores=0.5509009,
final_local_scores=0.5443621, pem=-0.13856095,
similarity=21.088081]
25: 14653 Iran
[pem_pos=0, final_scores=0.30370277,

```
final_local_scores=0.27433574, pem=-0.57066226,
similarity=15.379787]
26: 7515928 Iraq
    [pem_pos=0, final_scores=0.5980206,
final_local_scores=0.5514213, pem=-0.25379938,
similarity=21.925425]
27: 25008 Prisoner_of_war
    [pem_pos=0, final_scores=0.15913354,
final_local_scores=0.16830638, pem=0.0, similarity=9.112311]
```

Sample: b'APW19980917_0818_htm'

Cult 1 guru 's lawyers lash at prosecutors , court Attorneys for a 2 cult guru accused of murder in the 3 nerve gas attack on [1 1 Tokyo 1] subways denounced his trial Thursday as `` unfair '' because some of the prosecutors ' charges are contradicted by testimony from their own witnesses . [2 2 Shoko Asahara 2] , the leader of the [3 3 Aum Shinri Kyo 3] cult , is on trial on 17 charges in the March 20 , 1995 4 sar in attack that killed 12 people and sickened thousands . Other cult leaders have also been charged in the 5 gassing as well as in other slayings . Osamu Watanabe , [1 Asahara 1] 's chief 6 defense lawyer , said the trial should not be allowed to go on . `` We must wonder why such abnormal and unfair proceedings are tolerated . The court has been negligent , '' Watanabe told the [4 4 Tokyo District Court 4] . The prosecutors say [2 Asahara 2] , accused of masterminding the subway gassing , decided on the attack on March 18 , 1995 while riding in a car with his followers . But Watanabe pointed out that top cultist [3 Yoshihiro Inoue 3] , a key witness for the 7 prosecution , told the court that [4 Asahara 4] was still undecided on the timing of the attack at that time . Inoue quoted [5 Asahara 5] as saying that he needed to meditate about it . Last April , the defense raised similar objections , but the prosecutors refused to change their statement . -LRB- cw -RRB-

```
1: 30057 Tokyo
2: 98970 Shoko_Aсахara
3: 34267361 Aum_Shinrikyo
4: 6384526 Tokyo_District_Court
1: 167936 Guru
```

```
[winner: 168429 Guru_(rapper), pem_pos=1,
final_scores=-1.1320233, final_local_scores=-1.0873196,
pem=-0.9937921, similarity=-22.953676]
gt: 167936 Guru, pem_pos=0, final_scores=-1.2573907,
final_local_scores=-1.3020058, pem=-0.8709499,
similarity=-30.028376 ]
```

```
2: 1008588 Cult
[winner: 2175765 Ramtha's_School_of_Enlightenment,
pem_pos=0, final_scores=-1.2217611,
final_local_scores=-1.2583623, pem=-0.73966724,
similarity=-29.439905]
RECALL MISS: 1008588 Cult]
```


3: 21710 Nerve_agent
 [winner: 5170484 Infraorbital_nerve, pem_pos=0,
 final_scores=-0.627987, final_local_scores=-0.4560483,
 pem=0.0, similarity=-9.519678]
 gt: 21710 Nerve_agent, pem_pos=3,
 final_scores=-1.3563814, final_local_scores=-1.3903561,
 pem=-3.9280057, similarity=-16.040182]
 [winner: 18993869 Gas, pem_pos=0,
 final_scores=-1.1948683, final_local_scores=-1.1599164,
 pem=-0.3543779, similarity=-28.597345]
 RECALL MISS: 21710 Nerve_agent]

4: 448557 Sarin_gas_attack_on_the_Tokyo_subway
 [winner: 489711 Tokyo_subway, pem_pos=0,
 final_scores=-0.923938, final_local_scores=-0.91778874,
 pem=0.0, similarity=-23.298933]
 RECALL MISS: 448557 Sarin_gas_attack_on_the_Tokyo_subway]

5: 142788 Gas_chamber
 [winner: 22131 Natural_gas, pem_pos=1,
 final_scores=-1.3580592, final_local_scores=-1.3239958,
 pem=-0.5923425, similarity=-32.199715]
 gt: 142788 Gas_chamber, pem_pos=0,
 final_scores=-1.387408, final_local_scores=-1.4389622,
 pem=-0.32892498, similarity=-37.063046]

6: 228102 Defense_(legal)
 [winner: 228102 Defense_(legal), pem_pos=0,
 final_scores=-1.2309515, final_local_scores=-1.2748356,
 pem=-1.08081, similarity=-28.076313]
 gt: 228102 Defense_(legal), pem_pos=0,
 final_scores=-1.2309515, final_local_scores=-1.2748356,
 pem=-1.08081, similarity=-28.076313]

7: 440936 Prosecutor
 [winner: 92608 Witness_for_the_Prosecution_(1957_film),
 pem_pos=0, final_scores=-0.92439556,
 final_local_scores=-0.8415888, pem=0.0,
 similarity=-21.024975]
 RECALL MISS: 440936 Prosecutor]

1: 30057 Tokyo
 [pem_pos=0, final_scores=0.42122138,
 final_local_scores=0.4932196, pem=-0.09283106,
 similarity=19.3132]

2: 98970 Shoko_Asahara
 [pem_pos=0, final_scores=0.4361068,
 final_local_scores=0.39320368, pem=0.0, similarity=15.823694]

3: 34267361 Aum_Shinrikyo
 [pem_pos=0, final_scores=0.3940236,
 final_local_scores=0.3943001, pem=0.0, similarity=15.856414]

4: 6384526 Tokyo_District_Court
 [pem_pos=0, final_scores=0.081467554,
 final_local_scores=0.12687641, pem=0.0, similarity=7.875958]

1: 98970 Shoko_Asahara
 [pem_pos=0, final_scores=0.24022608,
 final_local_scores=0.17923376, pem=0.0, similarity=9.438406]

2: 98970 Shoko_Aсахara
 [pem_pos=0, final_scores=0.22751777,
 final_local_scores=0.16535184, pem=0.0, similarity=9.024141]
 3: 448557 Sarin_gas_attack_on_the_Tokyo_subway
 [pem_pos=0, final_scores=0.2902388,
 final_local_scores=0.29259196, pem=0.0, similarity=12.82124]
 4: 98970 Shoko_Aсахara
 [pem_pos=0, final_scores=0.32483077,
 final_local_scores=0.2716515, pem=0.0, similarity=12.196336]
 5: 98970 Shoko_Aсахara
 [pem_pos=0, final_scores=0.2899896,
 final_local_scores=0.23359281, pem=0.0, similarity=11.060588]

Sample: b'APW19980603_1617_htm'

[1 1 MALAYSIA 1] PORT DICKSON , Tues. - Federal and State 1 G
 overnment agencies , given the task to undertake 2 economic
 projects under the Development Programme for the Hardcore
 Poor -LRB- PPRT -RRB- , must ensure their successful 3 implem
 entation . 4 Rural Development Minister Datuk Annuar Musa
 said yesterday those responsible should closely monitor the
 implementation of the projects and ensure that the 5 allocati
 on provided by the 6 ministry benefit the hardcore 7 poor
 participants . He said those responsible , including district
 officers , must personally check on the progress of the
 projects being carried out . `` We do n't want to see
 projects as being a success on paper but on inspection , they
 do n't show the desired results , '' he added . According to
 him , it was the ministry 's policy now that the 8 minister ,
 deputy minister , parliamentary secretary , director-general
 and other senior ministry officials had to visit the PPRT
 projects to see what was happening to the allocation given by
 the Government . Annuar said this after visiting a 9 mussel
 rearing project in Kuala 10 Lukut today as part of his day-
 long visit to several PPRT projects in [2 2 Negri Sembilan 2]
 . Present were [1 Chembong 1] Assemblyman Firdaus Harun and [3
 3 Port Dickson 3] district officer Hassan Nawawi Abdul
 Rahman . Annuar added that the implementing agencies should
 also conduct in-depth 11 studies of the projects proposed to
 ensure their viability .

1: 3607937 Malaysia

2: 376437 Negeri_Sembilan

3: 1732134 Port_Dickson

1: 489674 Government_agency

[winner: 20629 Maharashtra, pem_pos=2,
 final_scores=-0.28742355, final_local_scores=-0.26150444,
 pem=-1.6071688, similarity=5.0259323]

RECALL MISS: 489674 Government_agency]

[winner: 32106147 Government_agencies_in_Iceland,
 pem_pos=4, final_scores=-1.2578033,
 final_local_scores=-1.1971352, pem=-2.6219954,
 similarity=-17.376373]

gt: 489674 Government_agency, pem_pos=15,

final_scores=-2.6222641, final_local_scores=-2.778038,
 pem=-6.9527287, similarity=-41.00246]
 2: 6639133 Economy
 [winner: 10822780 2006-
 07_economic_sanctions_against_the_Palestinian_National_Author
 ity, pem_pos=9, final_scores=-1.5749254,
 final_local_scores=-1.5754852, pem=-2.1972246,
 similarity=-30.97706|
 gt: 6639133 Economy, pem_pos=0, final_scores=-1.7966821,
 final_local_scores=-1.8792894, pem=-0.647923,
 similarity=-48.468525]
 3: 407704 Implementation
 [winner: 5439284 Programming_language_implementation,
 pem_pos=0, final_scores=-1.1648388,
 final_local_scores=-1.0940716, pem=0.0,
 similarity=-28.559568|
 gt: 407704 Implementation, pem_pos=1,
 final_scores=-1.4838685, final_local_scores=-1.4731193,
 pem=-0.3725879, similarity=-37.844913]
 4: 212614 Rural_area
 [winner: 8417961 Rural_development, pem_pos=0,
 final_scores=-0.1594517, final_local_scores=-0.095048964,
 pem=-0.20739101, similarity=2.3810868|
 RECALL MISS: 212614 Rural_area]
 5: 267428 Resource_allocation
 [winner: 485122 Register_allocation, pem_pos=1,
 final_scores=-1.3189565, final_local_scores=-1.2456417,
 pem=-0.93392575, similarity=-28.003885|
 gt: 267428 Resource_allocation, pem_pos=0,
 final_scores=-1.3228165, final_local_scores=-1.2966259,
 pem=0.0, similarity=-34.60419]
 6: 859401 Ministry_(government_department)
 [winner: 859412 Ministry_(collective_executive),
 pem_pos=0, final_scores=-0.75244886,
 final_local_scores=-0.67183983, pem=0.0,
 similarity=-15.959328|
 RECALL MISS: 859401 Ministry_(government_department)]
 7: 64959 Poverty
 [winner: 7324367 Act_for_the_Relief_of_the_Poor_1601,
 pem_pos=1, final_scores=-1.2578797,
 final_local_scores=-1.2457142, pem=-0.69234294,
 similarity=-29.319817|
 gt: 64959 Poverty, pem_pos=0, final_scores=-1.356613,
 final_local_scores=-1.382196, pem=-0.20411816,
 similarity=-36.047745]
 8: 739237 Minister_(government)
 [winner: 31835428 The_Minister, pem_pos=0,
 final_scores=-0.6916383, final_local_scores=-0.54149413,
 pem=0.0, similarity=-12.069551|
 RECALL MISS: 739237 Minister_(government)]
 9: 144156 Mussel
 [winner: 2572321 Perna_canalicula, pem_pos=0,

```

final_scores=-0.90028274, final_local_scores=-0.82674235,
pem=0.0, similarity=-20.581928|
    gt: 144156 Mussel, pem_pos=1, final_scores=-1.1059192,
final_local_scores=-1.0495238, pem=-0.006779713,
similarity=-27.1933 ]
10: 6127743 Lukut
    [winner: 6127743 Lukut, pem_pos=0,
final_scores=-0.6371979, final_local_scores=-0.6341851,
pem=0.0, similarity=-14.835634|
    gt: 6127743 Lukut, pem_pos=0, final_scores=-0.6371979,
final_local_scores=-0.6341851, pem=0.0,
similarity=-14.835634 ]
11: 25524 Research
    [winner: 12313415 Color_Marketing_Group, pem_pos=9,
final_scores=-1.3326561, final_local_scores=-1.2986734,
pem=-2.5455313, similarity=-20.822294|
    gt: 25524 Research, pem_pos=0, final_scores=-1.6596763,
final_local_scores=-1.7326672, pem=-1.4527498,
similarity=-39.71625 ]
1: 3607937 Malaysia
    [pem_pos=0, final_scores=0.16970463,
final_local_scores=0.16798443, pem=-0.20223404,
similarity=10.202483]
2: 376437 Negeri_Sembilan
    [pem_pos=0, final_scores=0.35833377,
final_local_scores=0.3506354, pem=0.0, similarity=14.553373]
3: 1732134 Port_Dickson
    [pem_pos=0, final_scores=0.20029782,
final_local_scores=0.22660479, pem=0.0, similarity=10.852052]
1: 3802282 Chembong
    [pem_pos=0, final_scores=0.12528796,
final_local_scores=0.22775182, pem=0.0, similarity=10.886282]

```

Sample: b'APW19980824_0827_hm'

[1 Israel 1] increases 1 security at airports following threats [1 1 Israel 1] is adding security personnel at its 2 international airport in response to threats by [2 3 Islamic 2] militants to attack [3 Israeli 3] and [2 2 U.S. 2] targets , including airlines , officials said Monday . Prime Minister [3 3 Benjamin Netanyahu 3] 's senior adviser , [4 4 David Bar-Illan 4] , said [4 Israel 4] would use all measures it deemed necessary to fight 4 terrorism . Asked whether this included 5 pre-emptive strikes , [5 Bar-Illan 5] told [6 The Associated Press 6] : `` I do n't want to be specific about anything but obviously we are not ruling out anything . '' Threats against [7 U.S. 7] and [8 Israeli 8] targets emerged after last week 's [9 U.S. 9] 6 air strikes in [5 5 Sudan 5] and [6 6 Afghani stan 6] that were aimed at 7 Osama bin Laden , a [7 7 Saudi 7] millionaire and [10 Islamic 10] militant linked to the bombings of two [11 U.S. 11] 8 embassies in [8 8 East Africa 8] this month . On Sunday , a leading [9 9 Muslim 9] 9 activist close to bin [12 Laden 12] said [13 Islamic 13] militants are

preparing to retaliate for the air strikes by targeting [14 U.S. 14] and [15 Israeli 15] 10 strategic sites and airliners . A [16 U.S. 16] basketball team from the [10 10 University of Connecticut 10] , which arrived Monday for a nine-day tour of friendly games , hired two private 11 security guards to accompany them in [17 Israel 17] . `` The 12 administration of the [18 University of Connecticut 18] demanded bodyguards and we readily agreed , '' said the teams ' 13 tour operator Nels Hawkinson . `` Had it not been for the bombings in [19 Afghanistan 19] and [20 Sudan 20] we would not have seen the need , '' he said . Pini Shis , 6th graf pvs dl

1: 9282173 Israel
2: 3434750 United_States
3: 56472 Benjamin_Netanyahu
4: 4727236 David_Bar-Illan
5: 27421 Sudan
6: 737 Afghanistan
7: 349303 Saudi_Arabia
8: 260987 East_Africa
9: 19541 Muslim
10: 239846 University_of_Connecticut
1: 41684 Security

[winner: 2593895 Protective_security_units, pem_pos=8, final_scores=-1.0294904, final_local_scores=-0.9901677, pem=-2.2968068, similarity=-12.9684725]
gt: 41684 Security, pem_pos=0, final_scores=-1.3000346, final_local_scores=-1.3957798, pem=-0.84670097, similarity=-32.958652]

2: 1519705 International_airport

[winner: 2154303
Licenciado_Gustavo_Díaz_Ordaz_International_Airport,
pem_pos=0, final_scores=-0.9556237,
final_local_scores=-0.8837129, pem=0.0,
similarity=-22.282042]

RECALL MISS: 1519705 International_airport]

3: 566821 Islamic_terrorism

[winner: 6037917 Islam, pem_pos=0, final_scores=0.35047835, final_local_scores=0.35355222, pem=-0.21141872, similarity=15.790145]
gt: 566821 Islamic_terrorism, pem_pos=2, final_scores=-0.19844669, final_local_scores=-0.22709927, pem=-2.6830015, similarity=11.903197]

[winner: 4643078 Conflict_in_the_Niger_Delta, pem_pos=0, final_scores=-1.3904445, final_local_scores=-1.4451833, pem=0.0, similarity=-39.03744]

gt: 566821 Islamic_terrorism, pem_pos=10, final_scores=-2.5395477, final_local_scores=-2.7867014, pem=-4.691348, similarity=-53.55874]

4: 30636 Terrorism

[winner: 30636 Terrorism, pem_pos=0, final_scores=-1.1100746, final_local_scores=-1.249903, pem=-0.18121491, similarity=-32.22441]

gt: 30636 Terrorism, pem_pos=0, final_scores=-1.1100746,
final_local_scores=-1.249903, pem=-0.18121491,
similarity=-32.22441]

5: 384984 Preemptive_war

[winner: 384984 Preemptive_war, pem_pos=0,
final_scores=-1.0392396, final_local_scores=-1.1018324,
pem=0.0, similarity=-28.791164]

gt: 384984 Preemptive_war, pem_pos=0,
final_scores=-1.0392396, final_local_scores=-1.1018324,
pem=0.0, similarity=-28.791164]

[winner: 574095 Strike_(attack), pem_pos=1,
final_scores=-1.3634723, final_local_scores=-1.3676395,
pem=-1.100119, similarity=-30.740763]

RECALL MISS: 384984 Preemptive_war]

6: 323167 Airstrike

[winner: 22282934 Marshalls-Gilberts_raids, pem_pos=0,
final_scores=-0.88292927, final_local_scores=-0.8215437,
pem=0.0, similarity=-20.42679]

gt: 323167 Airstrike, pem_pos=1,
final_scores=-1.1897267, final_local_scores=-1.2737477,
pem=-0.20340106, similarity=-32.815334]

7: 22468 Osama_bin_Laden

[winner: 22468 Osama_bin_Laden, pem_pos=1,
final_scores=0.024713203, final_local_scores=0.012293413,
pem=0.0, similarity=4.456573]

gt: 22468 Osama_bin_Laden, pem_pos=1,
final_scores=0.024713203, final_local_scores=0.012293413,
pem=0.0, similarity=4.456573]

[winner: 22468 Osama_bin_Laden, pem_pos=0,
final_scores=-0.004073739, final_local_scores=-0.026511788,
pem=-0.30829534, similarity=4.9751062]

gt: 22468 Osama_bin_Laden, pem_pos=0,
final_scores=-0.004073739, final_local_scores=-0.026511788,
pem=-0.30829534, similarity=4.9751062]

[winner: 287790 Filesystem_Hierarchy_Standard, pem_pos=4,
final_scores=-1.6888138, final_local_scores=-1.6696846,
pem=-2.6532419, similarity=-31.308264]

gt: 22468 Osama_bin_Laden, pem_pos=6,
final_scores=-2.9784138, final_local_scores=-3.2755265,
pem=-2.7586026, similarity=-78.65682]

8: 8970 Diplomatic_mission

[winner: 8970 Diplomatic_mission, pem_pos=0,
final_scores=-1.4101901, final_local_scores=-1.5523088,
pem=-0.030808141, similarity=-42.066742]

gt: 8970 Diplomatic_mission, pem_pos=0,
final_scores=-1.4101901, final_local_scores=-1.5523088,
pem=-0.030808141, similarity=-42.066742]

9: 39766702 Activism

[winner: 798781 Youth_activism, pem_pos=3,
final_scores=-1.6473833, final_local_scores=-1.7465911,
pem=-1.0986123, similarity=-42.05763]

gt: 39766702 Activism, pem_pos=0,

final_scores=-1.9634527, final_local_scores=-2.1662443,
pem=-0.1945611, similarity=-59.497284]
10: 29607 Strategy
[winner: 214209 Strategic_bomber, pem_pos=2,
final_scores=-1.6427276, final_local_scores=-1.7167597,
pem=-1.5961827, similarity=-38.461533]
gt: 29607 Strategy, pem_pos=0, final_scores=-1.702154,
final_local_scores=-1.8389034, pem=-0.25153232,
similarity=-49.418964]
11: 19283898 Security_guard
[winner: 2593895 Protective_security_units, pem_pos=8,
final_scores=-1.0330052, final_local_scores=-0.9940072,
pem=-2.2968068, similarity=-13.083051]
gt: 19283898 Security_guard, pem_pos=4,
final_scores=-1.6112411, final_local_scores=-1.7197719,
pem=-2.26688, similarity=-34.904064]
[winner: 18057247 Guard_Corps_(Haganah), pem_pos=17,
final_scores=-1.3552133, final_local_scores=-1.3862994,
pem=-3.0910425, similarity=-20.470657]
RECALL MISS: 19283898 Security_guard]
12: 1270090 Academic_administration
[winner: 361745 Presidency_of_Bill_Clinton, pem_pos=0,
final_scores=-1.1217116, final_local_scores=-1.2241709,
pem=0.0, similarity=-32.44199]
RECALL MISS: 1270090 Academic_administration]
13: 1114504 Tour_operator
[winner: 1114504 Tour_operator, pem_pos=0,
final_scores=-1.2671719, final_local_scores=-1.27985,
pem=-0.0048780614, similarity=-34.077034]
gt: 1114504 Tour_operator, pem_pos=0,
final_scores=-1.2671719, final_local_scores=-1.27985,
pem=-0.0048780614, similarity=-34.077034]
1: 9282173 Israel
[pem_pos=0, final_scores=0.63816494,
final_local_scores=0.6052112, pem=-0.21279564,
similarity=23.307638]
2: 3434750 United_States
[pem_pos=0, final_scores=0.5161851,
final_local_scores=0.48880988, pem=-0.14075686,
similarity=19.442234]
3: 56472 Benjamin_Netanyahu
[pem_pos=0, final_scores=0.62667525,
final_local_scores=0.6751363, pem=0.0, similarity=24.237127]
4: 4727236 David_Bar-Illan
[pem_pos=0, final_scores=0.29505944,
final_local_scores=0.40191245, pem=0.0, similarity=16.083582]
5: 27421 Sudan
[pem_pos=0, final_scores=0.4235403,
final_local_scores=0.40190065, pem=-0.2848764,
similarity=17.632433]
6: 737 Afghanistan
[pem_pos=0, final_scores=0.51922435,

final_local_scores=0.51529425, pem=-0.22286008,
similarity=20.67907]
7: 349303 Saudi_Arabia
[pem_pos=0, final_scores=0.6310569,
final_local_scores=0.6126231, pem=0.0, similarity=22.371609]
8: 260987 East_Africa
[pem_pos=0, final_scores=0.17468394,
final_local_scores=0.16798183, pem=-0.16404411,
similarity=9.994723]
9: 19541 Muslim
[pem_pos=0, final_scores=0.061820343,
final_local_scores=0.029279634, pem=-0.22736222,
similarity=6.199907]
10: 239846 University_of_Connecticut
[pem_pos=0, final_scores=0.27462715,
final_local_scores=0.33645657, pem=0.0, similarity=14.130248]
1: 9282173 Israel
[pem_pos=0, final_scores=0.61326164,
final_local_scores=0.5780081, pem=-0.21279564,
similarity=22.495842]
2: 6037917 Islam
[pem_pos=0, final_scores=0.35047835,
final_local_scores=0.35355222, pem=-0.21141872,
similarity=15.790145]
3: 9282173 Israel
[pem_pos=0, final_scores=0.5422022,
final_local_scores=0.5003864, pem=-0.34790626,
similarity=20.91421]
4: 9282173 Israel
[pem_pos=0, final_scores=0.681472,
final_local_scores=0.6525176, pem=-0.21279564,
similarity=24.719355]
5: 4727236 David_Bar-Illan
[pem_pos=0, final_scores=0.072032675,
final_local_scores=0.15828961, pem=0.0, similarity=8.813391]
6: 18935732 Associated_Press
[pem_pos=0, final_scores=0.3225981,
final_local_scores=0.34572196, pem=0.0, similarity=14.406746]
7: 3434750 United_States
[pem_pos=0, final_scores=0.55423886,
final_local_scores=0.5303778, pem=-0.14075686,
similarity=20.682701]
8: 9282173 Israel
[pem_pos=0, final_scores=0.53369886,
final_local_scores=0.49109787, pem=-0.34790626,
similarity=20.637022]
9: 3434750 United_States
[pem_pos=0, final_scores=0.50198287,
final_local_scores=0.4732961, pem=-0.14075686,
similarity=18.97927]
10: 6037917 Islam
[pem_pos=0, final_scores=0.3934908,


```

final_local_scores=0.40053678, pem=-0.21141872,
similarity=17.192257]
11: 3434750 United_States
    [pem_pos=0, final_scores=0.5316309,
final_local_scores=0.50568205, pem=-0.14075686,
similarity=19.945732]
12: 22468 Osama_bin_Laden
    [pem_pos=1, final_scores=0.22697668,
final_local_scores=0.2419829, pem=0.0, similarity=11.310966]
13: 6037917 Islam
    [pem_pos=0, final_scores=0.35604423,
final_local_scores=0.35963213, pem=-0.21141872,
similarity=15.9715805]
14: 3434750 United_States
    [pem_pos=0, final_scores=0.5568035,
final_local_scores=0.53317934, pem=-0.14075686,
similarity=20.766306]
15: 9282173 Israel
    [pem_pos=0, final_scores=0.5182082,
final_local_scores=0.4741767, pem=-0.34790626,
similarity=20.13206]
16: 3434750 United_States
    [pem_pos=0, final_scores=0.5804953,
final_local_scores=0.55905896, pem=-0.14075686,
similarity=21.538605]
17: 9282173 Israel
    [pem_pos=0, final_scores=0.5473564,
final_local_scores=0.5060167, pem=-0.21279564,
similarity=20.347477]
18: 239846 University_of_Connecticut
    [pem_pos=0, final_scores=0.32358068,
final_local_scores=0.38993084, pem=0.0, similarity=15.726027]
19: 737 Afghanistan
    [pem_pos=0, final_scores=0.559959,
final_local_scores=0.5597906, pem=-0.22286008,
similarity=22.00693]
20: 27421 Sudan
    [pem_pos=0, final_scores=0.5303316,
final_local_scores=0.518554, pem=-0.2848764,
similarity=21.1136]

```

Sample: b'APW19981113_0500_htm'

Airport ready for operation ; [1 Palestinians 1] await word from [2 Israel 2] The new [3 1 Palestinian 3] 2 airport , a powerful symbol of fledgling independence , is ready to open , the head of the Palestinian 3 Civil Aviation Authority said Friday after an operations drill with [4 4 Israeli 4] officials . [1 1 Yasser Arafat International Airport 1] in the southern [2 2 Gaza Strip 2] was to have opened last week as part of the Israeli-Palestinian peace agreement . However , [5 Israel 5] fell behind schedule when it held up Cabinet ratification of the accord by two weeks . Prime

Minister [6 Benjamin Netanyahu 6] has said the airport would start operating next week , but has not set a date . On Thursday , [7 Israeli 7] and [8 Palestinian 8] officials staged a four-hour drill , said Fayez Zeidan , head of the Palestinian 5 Civil Aviation Authority . About 100 [9 Palestinian 9] airport employees and 50 [10 6 Israelis 10] who will share 7 security duties played the roles of passengers and airport personnel . For lack of a plane , the `` passengers `` arrived by 8 bus . Zeidan said the exercise covered arrival and departure procedures , including passport checks as well as customs and luggage inspection . He said the airport was ready for operation . `` We just need to hear from the [11 Israelis 11] , `` Zeidan said . sa-kl

1: 40390174 Yasser_Arafat_International_Airport

2: 12047 Gaza_Strip

1: 23267 Palestinian_people

[winner: 241405 State_of_Palestine, pem_pos=9, final_scores=0.32954425, final_local_scores=0.27903658, pem=-2.8508017, similarity=27.919825]

gt: 23267 Palestinian_people, pem_pos=0, final_scores=-0.3484519, final_local_scores=-0.45833632, pem=-0.38816506, similarity=-7.4770546]

2: 37575 Airport

[winner: 37575 Airport, pem_pos=0, final_scores=-1.349098, final_local_scores=-1.3487384, pem=-0.121571906, similarity=-35.498207]

gt: 37575 Airport, pem_pos=0, final_scores=-1.349098, final_local_scores=-1.3487384, pem=-0.121571906, similarity=-35.498207]

3: 18575446 Civil_aviation

[winner: 20778386 National_aviation_authority, pem_pos=0, final_scores=-0.56267, final_local_scores=-0.53579104, pem=-0.40803733, similarity=-9.680391]

RECALL MISS: 18575446 Civil_aviation]

4: 181296 Israelis

[winner: 9282173 Israel, pem_pos=0, final_scores=0.5646126, final_local_scores=0.5109461, pem=-0.34790626, similarity=21.22933]

gt: 181296 Israelis, pem_pos=1, final_scores=-0.35006, final_local_scores=-0.40807098, pem=-1.7488571, similarity=1.4226176]

5: 20778386 National_aviation_authority

[winner: 684737

Civil_Aviation_Authority_(United_Kingdom), pem_pos=1, final_scores=-0.60362417, final_local_scores=-0.5756376, pem=-0.8439734, similarity=-8.498801]

gt: 20778386 National_aviation_authority, pem_pos=0, final_scores=-0.65042514, final_local_scores=-0.6316502, pem=-0.40803733, similarity=-12.5410185]

6: 181296 Israelis

[winner: 9282173 Israel, pem_pos=1, final_scores=0.3196137, final_local_scores=0.24332199,

pem=-0.5460739, similarity=14.320561|
gt: 181296 Israelis, pem_pos=0,
final_scores=-0.0048510134, final_local_scores=-0.030982628,
pem=-0.009432555, similarity=3.2164252]
7: 41684 Security
[winner: 2593895 Protective_security_units, pem_pos=8,
final_scores=-0.8940158, final_local_scores=-0.8608815,
pem=-2.2968068, similarity=-9.110317|
gt: 41684 Security, pem_pos=0, final_scores=-1.1713604,
final_local_scores=-1.2377081, pem=-0.84670097,
similarity=-28.241472]
8: 4146 Bus
[winner: 21601728 Trolleybuses_in_Quito, pem_pos=0,
final_scores=-1.1513779, final_local_scores=-1.0890695,
pem=-0.65058756, similarity=-24.872297|
RECALL MISS: 4146 Bus]
1: 40390174 Yasser_Arafat_International_Airport
[pem_pos=0, final_scores=0.16708405,
final_local_scores=0.22800577, pem=0.0, similarity=10.89386]
2: 12047 Gaza_Strip
[pem_pos=0, final_scores=0.39627331,
final_local_scores=0.3663777, pem=0.0, similarity=15.023154]
1: 241405 State_of_Palestine
[pem_pos=5, final_scores=0.27951074,
final_local_scores=0.22438258, pem=-1.7909594,
similarity=20.525253]
2: 9282173 Israel
[pem_pos=0, final_scores=0.5139814,
final_local_scores=0.45563918, pem=-0.21279564,
similarity=18.844112]
3: 241405 State_of_Palestine
[pem_pos=9, final_scores=0.32954425,
final_local_scores=0.27903658, pem=-2.8508017,
similarity=27.919825]
4: 9282173 Israel
[pem_pos=0, final_scores=0.5646126,
final_local_scores=0.5109461, pem=-0.34790626,
similarity=21.22933]
5: 9282173 Israel
[pem_pos=0, final_scores=0.6453898,
final_local_scores=0.5991829, pem=-0.21279564,
similarity=23.127743]
6: 56472 Benjamin_Netanyahu
[pem_pos=0, final_scores=0.6651466,
final_local_scores=0.6864976, pem=0.0, similarity=24.576172]
7: 9282173 Israel
[pem_pos=0, final_scores=0.5411491,
final_local_scores=0.48531574, pem=-0.34790626,
similarity=20.46447]
8: 241405 State_of_Palestine
[pem_pos=9, final_scores=0.67554766,
final_local_scores=0.6569928, pem=-2.8508017,

```

similarity=39.19879]
9: 241405 State_of_Palestine
   [pem_pos=9, final_scores=0.3036512,
final_local_scores=0.2507524, pem=-2.8508017,
similarity=27.075768]
10: 9282173 Israel
    [pem_pos=1, final_scores=0.3196137,
final_local_scores=0.24332199, pem=-0.5460739,
similarity=14.320561]
11: 9282173 Israel
    [pem_pos=1, final_scores=0.29046983,
final_local_scores=0.21148673, pem=-0.5460739,
similarity=13.370534]

```

Sample: b'APW19980604_0787_htm'

ENGLISH 1 Viravaidya 's plan to set up a watchdog company to monitor 2 SET - listed companies and expose possible abuses of 3 corporate ethics . I agree with him whole-heartedly and would like to know where to mail my checks to subscribe to the company 's shares . Could you advise ? Also , I fully support his request for the media to report on the 4 lifestyle of former 5 executives of the failed 6 banks and 7 finance companies . We all know too well that these people will not be punished legally because of the inefficiency of the [1 1 Bank of Thailand 1] in collecting evidence against them and of our justice system in delivering timely justice . It is the media 's job to discover how these guys are better off after destroying our country , and it is the [1 8 Thai 1] people 's job to socially punish them . I would like to propose that 9 newspapers and 10 TV stations play up this issue -LRB- the continued luxurious lifestyle of the former executives -RRB- to attract more public attention , and I will give them Bt20 , 000 of my hard-earned salary with my full blessing . Artit Krichphiphat -LRB- An ordinary [2 Thai 2] citizen who wants to see the right things get done -RRB- Editor 's note : Khun Mechai will organise a seminar on the plan on June 7 at the Telephone Organisation of Thailand at 10 am . The media already has the duty to cover such issues without additional financial reward .

1: 749269 Bank_of_Thailand

1: 834384 Mechai_Viravaidya

```

   [winner: 834384 Mechai_Viravaidya, pem_pos=0,
final_scores=-8.057058e-05, final_local_scores=0.07614648,
pem=0.0, similarity=6.362076]

```

```

   gt: 834384 Mechai_Viravaidya, pem_pos=0,
final_scores=-8.057058e-05, final_local_scores=0.07614648,
pem=0.0, similarity=6.362076 ]

```

2: 7307109 SET_Index

```

   [winner: 59236 Set_(mythology), pem_pos=0,
final_scores=-0.45212168, final_local_scores=-0.35187465,
pem=-0.49313, similarity=-3.7292106]

```

```

   gt: 7307109 SET_Index, pem_pos=17,

```

final_scores=-0.771132, final_local_scores=-0.7343848,
pem=-3.4965076, similarity=1.188759]
3: 4770 Business_ethics
[winner: 9285 Ethical_naturalism, pem_pos=3,
final_scores=-0.96375215, final_local_scores=-0.86801153,
pem=-1.3860223, similarity=-14.276077]
gt: 4770 Business_ethics, pem_pos=9,
final_scores=-1.8848057, final_local_scores=-1.9630046,
pem=-5.291742, similarity=-25.712946]
[winner: 7485 Corporation, pem_pos=0,
final_scores=-1.2627469, final_local_scores=-1.330595,
pem=-0.3120966, similarity=-33.920662]
RECALL MISS: 4770 Business_ethics]
4: 190041 Lifestyle_(sociology)
[winner: 19851116 Lifestyle_magazine, pem_pos=1,
final_scores=-1.2730156, final_local_scores=-1.2725229,
pem=-1.3222258, similarity=-26.69444]
gt: 190041 Lifestyle_(sociology), pem_pos=0,
final_scores=-1.4424609, final_local_scores=-1.5028214,
pem=-0.36584687, similarity=-38.767944]
5: 52234 Chief_executive_officer
[winner: 12921466 Charbel_Nader, pem_pos=4,
final_scores=-0.97979236, final_local_scores=-0.9115275,
pem=-1.609438, similarity=-14.359713]
gt: 52234 Chief_executive_officer, pem_pos=9,
final_scores=-2.1473799, final_local_scores=-2.2921288,
pem=-4.8152285, similarity=-38.126022]
6: 19360669 Bank
[winner: 24691130 Bank_(card_game), pem_pos=4,
final_scores=-1.1582613, final_local_scores=-1.0525866,
pem=-1.3858787, similarity=-19.784945]
gt: 19360669 Bank, pem_pos=0, final_scores=-1.5544188,
final_local_scores=-1.6626523, pem=-0.087153114,
similarity=-45.053192]
7: 11162 Finance
[winner: 261584 Department_of_Finance_Canada, pem_pos=3,
final_scores=-1.3093157, final_local_scores=-1.322159,
pem=-1.7913799, similarity=-25.624352]
gt: 11162 Finance, pem_pos=0, final_scores=-1.5471468,
final_local_scores=-1.650326, pem=-0.045090236,
similarity=-44.9141]
8: 2670504 Thai_people
[winner: 30128 Thailand, pem_pos=2,
final_scores=0.5632963, final_local_scores=0.57836515,
pem=-1.7504287, similarity=30.868385]
gt: 2670504 Thai_people, pem_pos=3,
final_scores=-0.28279155, final_local_scores=-0.31575742,
pem=-1.801677, similarity=4.4646807]
[winner: 14654277 The_Autumn_People, pem_pos=17,
final_scores=-1.4139993, final_local_scores=-1.3235083,
pem=-3.2580965, similarity=-17.68838]
RECALL MISS: 2670504 Thai_people]

9: 22112840 Newspaper
 [winner: 12740212 Tab_Communications, pem_pos=1,
 final_scores=-1.1123475, final_local_scores=-1.0592287,
 pem=-0.6931472, similarity=-23.75034]
 gt: 22112840 Newspaper, pem_pos=0,
 final_scores=-1.2397214, final_local_scores=-1.277925,
 pem=-0.015627565, similarity=-33.961132]
 10: 29840 Television_channel
 [winner: 29831 Television, pem_pos=0,
 final_scores=-0.39558023, final_local_scores=-0.36674443,
 pem=-0.13274649, similarity=-6.1327763]
 RECALL MISS: 29840 Television_channel]
 [winner: 15256577 Rowley_(MBTA_station), pem_pos=8,
 final_scores=-1.63761, final_local_scores=-1.5855243,
 pem=-2.50274, similarity=-29.61521]
 RECALL MISS: 29840 Television_channel]
 1: 749269 Bank_of_Thailand
 [pem_pos=0, final_scores=0.53037506,
 final_local_scores=0.6448587, pem=0.0, similarity=23.333584]
 1: 30128 Thailand
 [pem_pos=2, final_scores=0.5632963,
 final_local_scores=0.57836515, pem=-1.7504287,
 similarity=30.868385]
 2: 30128 Thailand
 [pem_pos=2, final_scores=0.46297866,
 final_local_scores=0.46878332, pem=-1.7504287,
 similarity=27.598246]

Sample: b'APW19981109_0440_hm'

REPEATING Protectively [1 1 Iran 1] will protest to the [1
 International Court of Justice 1] at 2 the [2 Hague 2] and
 other global bodies about the U.S.-funded [3 3 Radio Free Eur
 ope 3] , the Iran Daily reported Monday . It quoted Foreign
 Minister [4 4 Kamal Kharrazi 4] as saying the radio `` was
 set up to interfere in [2 Iran 2] 's internal affairs . '' It
 did not say when the complaints will be filed . The [3
 English-language 3] daily also did not say which
 organizations [4 Iran 4] will approach besides the [5 5 ICJ 5]
 in 6 the [6 Netherlands 6] . On Tuesday , [5 Iran 5] recalled
 its 1 ambassador from [7 7 Prague 7] to protest the [8 8 Czech
 8] government 's assistance to the radio and said it would
 reduce the level of its economic and political cooperation
 with [6 Prague 6] . The American-sponsored station also began
 2 broadcast s directed to [9 9 Iraq 9] on Friday . In a trial
 period of several weeks , the station will broadcast one 30-
 minute program a day to [7 Iran 7] and [8 Iraq 8] . The [10 10
 Farsi 10] - language service to [9 Iran 9] was approved by
 the [10 3 Czech 10] government in August . [11 Radio Free
 Europe 11] began transmitting from [11 11 Munich 11] , [12 Ge
 rmany 12] , in 1951 , spreading uncensored news to [12 12 Sov
 iet 12] - controlled countries behind the [13 13 Iron Curtain
 13] during the [14 14 Cold War 14] between the West and the [1

3 4 Communist East 13] . It moved its headquarters to [14 Prague 14] in 1995 following the 5 collapse of communism six years earlier . ti/vj

- 1: 14653 Iran
- 2: 30269 The_Hague
- 3: 184408 Radio_Free_Europe/Radio_Liberty
- 4: 186801 Kamal_Kharazi
- 5: 14918 International_Court_of_Justice
- 6: 21148 Netherlands
- 7: 23844 Prague
- 8: 5321 Czech_Republic
- 9: 7515928 Iraq
- 10: 11600 Persian_language
- 11: 19058 Munich
- 12: 26779 Soviet_Union
- 13: 18953051 Iron_Curtain
- 14: 325329 Cold_War

1: 68651 Ambassador

[winner: 68651 Ambassador, pem_pos=0,
final_scores=-1.3223497, final_local_scores=-1.4608622,
pem=-0.027137717, similarity=-39.357746|

gt: 68651 Ambassador, pem_pos=0,
final_scores=-1.3223497, final_local_scores=-1.4608622,
pem=-0.027137717, similarity=-39.357746]

2: 113604 Broadcasting

[winner: 2704540 Broadcasting_(networking), pem_pos=7,
final_scores=-1.5292338, final_local_scores=-1.4918115,
pem=-2.2149708, similarity=-28.38356|

gt: 113604 Broadcasting, pem_pos=0,
final_scores=-1.6417481, final_local_scores=-1.7703571,
pem=-0.20117335, similarity=-47.647263]

3: 5848 Politics_of_the_Czech_Republic

[winner: 5321 Czech_Republic, pem_pos=1,
final_scores=0.5158078, final_local_scores=0.51294273,
pem=-1.208016, similarity=25.966326|

RECALL MISS: 5848 Politics_of_the_Czech_Republic]

[winner: 12229 Government, pem_pos=0,
final_scores=-1.5757508, final_local_scores=-1.7232585,
pem=-0.3899827, similarity=-45.21497|

RECALL MISS: 5848 Politics_of_the_Czech_Republic]

4: 9209651 Communism

[winner: 97477 Eastern_Bloc, pem_pos=0,
final_scores=0.18246882, final_local_scores=0.1676498,
pem=-0.6931472, similarity=12.86216|

RECALL MISS: 9209651 Communism]

5: 4584893 Revolutions_of_1989

[winner: 21352617 Galloping_Gertie, pem_pos=0,
final_scores=-0.9296781, final_local_scores=-0.8089165,
pem=-0.4054651, similarity=-17.844988|

RECALL MISS: 4584893 Revolutions_of_1989]

[winner: 9209651 Communism, pem_pos=0,
final_scores=-1.0203575, final_local_scores=-1.1500657,

pem=-0.07914143, similarity=-29.80016|
RECALL MISS: 4584893 Revolutions_of_1989]

- 1: 14653 Iran
[pem_pos=0, final_scores=0.6431096,
final_local_scores=0.6937983, pem=-0.13856095,
similarity=25.547554]
- 2: 30269 The_Hague
[pem_pos=0, final_scores=0.3281138,
final_local_scores=0.32778215, pem=-0.55011845,
similarity=16.863016]
- 3: 184408 Radio_Free_Europe/Radio_Liberty
[pem_pos=0, final_scores=0.41764134,
final_local_scores=0.43332767, pem=0.0, similarity=17.021074]
- 4: 186801 Kamal_Kharazi
[pem_pos=0, final_scores=0.38918686,
final_local_scores=0.49500126, pem=0.0, similarity=18.861538]
- 5: 14918 International_Court_of_Justice
[pem_pos=0, final_scores=0.1746035,
final_local_scores=0.1938951, pem=0.0, similarity=9.87593]
- 6: 21148 Netherlands
[pem_pos=0, final_scores=0.45301008,
final_local_scores=0.45471627, pem=-0.24940795,
similarity=19.015673]
- 7: 23844 Prague
[pem_pos=0, final_scores=0.4399343,
final_local_scores=0.4490561, pem=-0.05392049,
similarity=17.78367]
- 8: 5321 Czech_Republic
[pem_pos=1, final_scores=0.4878251,
final_local_scores=0.48237586, pem=-1.208016,
similarity=25.05415]
- 9: 7515928 Iraq
[pem_pos=0, final_scores=0.55079377,
final_local_scores=0.5491772, pem=-0.25379938,
similarity=21.858458]
- 10: 11600 Persian_language
[pem_pos=0, final_scores=0.054759547,
final_local_scores=0.04587476, pem=-0.17037486,
similarity=6.385232]
- 11: 19058 Munich
[pem_pos=0, final_scores=0.6122921,
final_local_scores=0.6358223, pem=-0.119245164,
similarity=23.712393]
- 12: 26779 Soviet_Union
[pem_pos=0, final_scores=0.31083876,
final_local_scores=0.2805608, pem=-0.097171314,
similarity=12.990639]
- 13: 18953051 Iron_Curtain
[pem_pos=0, final_scores=0.21661179,
final_local_scores=0.20137325, pem=0.0, similarity=10.0990925]
- 14: 325329 Cold_War
[pem_pos=0, final_scores=0.2478547,

final_local_scores=0.24821883, pem=0.0, similarity=11.497058]
1: 14918 International_Court_of_Justice
[pem_pos=0, final_scores=0.25715166,
final_local_scores=0.28406635, pem=0.0, similarity=12.566819]
2: 14653 Iran
[pem_pos=0, final_scores=0.6338981,
final_local_scores=0.6837362, pem=-0.13856095,
similarity=25.247282]
3: 8569916 English_language
[pem_pos=0, final_scores=0.22633828,
final_local_scores=0.2404345, pem=0.0, similarity=11.264758]
4: 14653 Iran
[pem_pos=0, final_scores=0.64125746,
final_local_scores=0.6917752, pem=-0.13856095,
similarity=25.48718]
5: 14653 Iran
[pem_pos=0, final_scores=0.5452507,
final_local_scores=0.58690244, pem=-0.13856095,
similarity=22.357569]
6: 23844 Prague
[pem_pos=0, final_scores=0.08368309,
final_local_scores=0.059905797, pem=-0.05392049,
similarity=6.1706495]
7: 14653 Iran
[pem_pos=0, final_scores=0.5361287,
final_local_scores=0.576938, pem=-0.13856095,
similarity=22.06021]
8: 7515928 Iraq
[pem_pos=0, final_scores=0.56387955,
final_local_scores=0.56347144, pem=-0.25379938,
similarity=22.285025]
9: 14653 Iran
[pem_pos=0, final_scores=0.58082706,
final_local_scores=0.6257642, pem=-0.13856095,
similarity=23.51728]
10: 5321 Czech_Republic
[pem_pos=1, final_scores=0.5158078,
final_local_scores=0.51294273, pem=-1.208016,
similarity=25.966326]
11: 184408 Radio_Free_Europe/Radio_Liberty
[pem_pos=0, final_scores=0.4574104,
final_local_scores=0.47676933, pem=0.0, similarity=18.317461]
12: 11867 Germany
[pem_pos=0, final_scores=0.8176643,
final_local_scores=0.8302581, pem=-0.15529114,
similarity=29.710768]
13: 97477 Eastern_Bloc
[pem_pos=0, final_scores=0.18246882,
final_local_scores=0.1676498, pem=-0.6931472,
similarity=12.86216]
14: 23844 Prague
[pem_pos=0, final_scores=0.36992633,

final_local_scores=0.37258303, pem=-0.05392049,
similarity=15.501562]

Sample: b'APW19981109_0140_htm'

[1 Iran 1] to complain about U.S.-radio at the [2 Hague 2] tribunal [3 Iran 3] will protest to the [4 International Court of Justice 4] at 1 the [1 Hague 1] and other global bodies about the U.S.-funded [2 2 Radio Free Europe 2], the Iran Daily reported Monday. It quoted Foreign Minister [3 3 Kamal Kharrazi 3] as saying the radio ``was set up to interfere in [5 Iran 5]'s internal affairs. '' It did not say when the complaints will be filed. The [6 English-language 6] daily also did not say which organizations [7 Iran 7] will approach besides the [4 4 ICJ 4] in 5 the [5 Netherlands 5]. On Tuesday, [8 Iran 8] recalled its 1 ambassador from [6 6 Prague 6] to protest the [7 7 Czech 7] government's assistance to the radio and said it would reduce the level of its 2 economic and 3 political cooperation with [9 Prague 9]. The American-sponsored station also began broadcasts directed to [8 8 Iraq 8] on Friday. In a trial period of several weeks, the station will broadcast one 30-minute program a day to [10 Iran 10] and [11 Iraq 11]. The [9 9 Farsi 9] - language service to [12 Iran 12] was approved by the [13 Czech 13] government in August. [14 Radio Free Europe 14] began transmitting from [10 10 Munich 10], [15 Germany 15], in 1951, spreading uncensored news to [11 11 Soviet 11] - controlled countries behind the [12 12 Iron Curtain 12] during the [13 13 Cold War 13] between the 5 West and the [16 6 Communist 7 East 16]. It moved its headquarters to [17 Prague 17] in 1995 following the collapse of communism six years earlier. ti/vj

1: 30269 The_Hague

2: 184408 Radio_Free_Europe/Radio_Liberty

3: 186801 Kamal_Kharrazi

4: 14918 International_Court_of_Justice

5: 21148 Netherlands

6: 23844 Prague

7: 5321 Czech_Republic

8: 7515928 Iraq

9: 11600 Persian_language

10: 19058 Munich

11: 26779 Soviet_Union

12: 18953051 Iron_Curtain

13: 325329 Cold_War

1: 68651 Ambassador

[winner: 68651 Ambassador, pem_pos=0,
final_scores=-1.3213556, final_local_scores=-1.4609067,
pem=-0.027137717, similarity=-39.359077]

gt: 68651 Ambassador, pem_pos=0,
final_scores=-1.3213556, final_local_scores=-1.4609067,
pem=-0.027137717, similarity=-39.359077]

2: 6639133 Economy

[winner: 57001 Economy_of_São_Tomé_and_Príncipe,
pem_pos=0, final_scores=-0.97850704,
final_local_scores=-0.94954014, pem=0.0,
similarity=-24.246458]
RECALL MISS: 6639133 Economy]

3: 22986 Politics
[winner: 24853 Political_media, pem_pos=11,
final_scores=-1.5777888, final_local_scores=-1.6041579,
pem=-2.6390574, similarity=-29.429955]
gt: 22986 Politics, pem_pos=0, final_scores=-1.8977058,
final_local_scores=-2.0894392, pem=-0.38701898,
similarity=-56.158653]

4: 17253537 Censorship
[winner: 5918986 Censorship_in_Thailand, pem_pos=2,
final_scores=-1.257435, final_local_scores=-1.2908692,
pem=-0.6743347, similarity=-30.765266]
gt: 17253537 Censorship, pem_pos=0,
final_scores=-1.3874234, final_local_scores=-1.5106614,
pem=-0.26186466, similarity=-39.567375]

5: 21208200 Western_world
[winner: 21208200 Western_world, pem_pos=0,
final_scores=-0.10400601, final_local_scores=-0.14145762,
pem=-1.6371275, similarity=8.771289]
gt: 21208200 Western_world, pem_pos=0,
final_scores=-0.10400601, final_local_scores=-0.14145762,
pem=-1.6371275, similarity=8.771289]

6: 9209651 Communism
[winner: 97477 Eastern_Bloc, pem_pos=0,
final_scores=0.17997335, final_local_scores=0.16819078,
pem=-0.6931472, similarity=12.878304]
RECALL MISS: 9209651 Communism]

7: 18984987 Eastern_world
[winner: 97477 Eastern_Bloc, pem_pos=0,
final_scores=0.17997335, final_local_scores=0.16819078,
pem=-0.6931472, similarity=12.878304]
RECALL MISS: 18984987 Eastern_world]

1: 30269 The_Hague
[pem_pos=0, final_scores=0.33145893,
final_local_scores=0.32762563, pem=-0.55011845,
similarity=16.858345]

2: 184408 Radio_Free_Europe/Radio_Liberty
[pem_pos=0, final_scores=0.4163291,
final_local_scores=0.43369323, pem=0.0, similarity=17.031984]

3: 186801 Kamal_Kharazi
[pem_pos=0, final_scores=0.39155924,
final_local_scores=0.4967608, pem=0.0, similarity=18.914045]

4: 14918 International_Court_of_Justice
[pem_pos=0, final_scores=0.17552315,
final_local_scores=0.19401786, pem=0.0, similarity=9.879593]

5: 21148 Netherlands
[pem_pos=0, final_scores=0.4541431,
final_local_scores=0.4546792, pem=-0.24940795,

similarity=19.014566]
6: 23844 Prague
[pem_pos=0, final_scores=0.43816704,
final_local_scores=0.44904482, pem=-0.05392049,
similarity=17.783333]
7: 5321 Czech_Republic
[pem_pos=1, final_scores=0.4860764,
final_local_scores=0.48239684, pem=-1.208016,
similarity=25.054775]
8: 7515928 Iraq
[pem_pos=0, final_scores=0.5396965,
final_local_scores=0.53700536, pem=-0.25379938,
similarity=21.495224]
9: 11600 Persian_language
[pem_pos=0, final_scores=0.05507289,
final_local_scores=0.0457986, pem=-0.17037486,
similarity=6.3829594]
10: 19058 Munich
[pem_pos=0, final_scores=0.60998756,
final_local_scores=0.6335058, pem=-0.119245164,
similarity=23.643265]
11: 26779 Soviet_Union
[pem_pos=0, final_scores=0.29830855,
final_local_scores=0.2697278, pem=-0.097171314,
similarity=12.66736]
12: 18953051 Iron_Curtain
[pem_pos=0, final_scores=0.22051327,
final_local_scores=0.20858628, pem=0.0, similarity=10.314343]
13: 325329 Cold_War
[pem_pos=0, final_scores=0.24683316,
final_local_scores=0.24948299, pem=0.0, similarity=11.534783]
1: 14653 Iran
[pem_pos=0, final_scores=0.5287046,
final_local_scores=0.5635387, pem=-0.13856095,
similarity=21.660347]
2: 30269 The_Hague
[pem_pos=0, final_scores=0.23310442,
final_local_scores=0.22018829, pem=-0.55011845,
similarity=13.652201]
3: 14653 Iran
[pem_pos=0, final_scores=0.6474037,
final_local_scores=0.69319946, pem=-0.13856095,
similarity=25.529684]
4: 14918 International_Court_of_Justice
[pem_pos=0, final_scores=0.25960165,
final_local_scores=0.28586084, pem=0.0, similarity=12.620371]
5: 14653 Iran
[pem_pos=0, final_scores=0.63865894,
final_local_scores=0.6836471, pem=-0.13856095,
similarity=25.244621]
6: 8569916 English_language
[pem_pos=0, final_scores=0.22582172,

final_local_scores=0.24041617, pem=0.0, similarity=11.264211]
7: 14653 Iran
[pem_pos=0, final_scores=0.6461426,
final_local_scores=0.6918219, pem=-0.13856095,
similarity=25.488573]
8: 14653 Iran
[pem_pos=0, final_scores=0.550127,
final_local_scores=0.58693945, pem=-0.13856095,
similarity=22.358673]
9: 23844 Prague
[pem_pos=0, final_scores=0.07666199,
final_local_scores=0.05415553, pem=-0.05392049,
similarity=5.99905]
10: 14653 Iran
[pem_pos=0, final_scores=0.5406614,
final_local_scores=0.5765997, pem=-0.13856095,
similarity=22.050116]
11: 7515928 Iraq
[pem_pos=0, final_scores=0.5633632,
final_local_scores=0.5628576, pem=-0.25379938,
similarity=22.266708]
12: 14653 Iran
[pem_pos=0, final_scores=0.58557165,
final_local_scores=0.6256573, pem=-0.13856095,
similarity=23.514091]
13: 5321 Czech_Republic
[pem_pos=1, final_scores=0.51421803,
final_local_scores=0.5131373, pem=-1.208016,
similarity=25.972132]
14: 184408 Radio_Free_Europe/Radio_Liberty
[pem_pos=0, final_scores=0.45804757,
final_local_scores=0.47926438, pem=0.0, similarity=18.391918]
15: 11867 Germany
[pem_pos=0, final_scores=0.81244713,
final_local_scores=0.82561636, pem=-0.15529114,
similarity=29.57225]
16: 97477 Eastern_Bloc
[pem_pos=0, final_scores=0.17997335,
final_local_scores=0.16819078, pem=-0.6931472,
similarity=12.878304]
17: 23844 Prague
[pem_pos=0, final_scores=0.3667869,
final_local_scores=0.37107295, pem=-0.05392049,
similarity=15.456497]

Sample: b'APW19980610_0111_htm'

Businessman killed in [1 Siberian 1] town A [2 1 Russian 2]
businessman was killed on his way to work Wednesday by an
unidentified gunman in what appeared to be an 2 organized cri
me killing , according to news reports . Yuri Zinin , the
head of 3 Sibneft Company was shot in the back as he walked
out of his apartment building in the town of Leninsk-

Kuzntesky in the [1 1 Kemerovo 1] region of central [2 2 Siberia 2] , [3 Russian 3] news agencies quoted local police as saying . Zinin 's company has the same name as one of [3 3 Russia 3] 's largest oil producers , [4 Sibneft 4] , controlled by controversial business 4 mogul [5 5 Boris Berezovsky . 5] But the two companies are separate . Zinin 's [6 Sibneft 6] was locally-based and mostly involved in selling 6 oil products and 7 coal . [7 Kemerovo 7] is one of [8 Russia 8] 's main 8 coal-mining regions . Police launched a search for the killer who fled . Officials would n't comment on possible motives for the murder . There are hundreds of 9 contract murders in [9 Russia 9] every year as 10 mobsters fight for control of profitable businesses . Only a small number of such killings are ever solved . The small town of 11 Leninsk-Kuznetsky became known nationwide last year , when the city 12 mayor was arrested on 13 tax fraud and 14 embezzlement charges . Mayor [4 4 Gennady Konyakhin 4] , also accused of being linked to the 15 mob , is in custody awaiting 16 trial . -LRB- vi/ren -RRB-

1: 17131 Kemerovo

2: 29639 Siberia

3: 25391 Russia

4: 2265916 Gennady_Konyakhin

1: 59510 Russians

[winner: 25391 Russia, pem_pos=1, final_scores=0.8221814, final_local_scores=0.80407524, pem=-1.0515022, similarity=33.803154]

gt: 59510 Russians, pem_pos=2, final_scores=-0.650902, final_local_scores=-0.75249356, pem=-2.236908, similarity=-6.201544]

2: 22625 Organized_crime

[winner: 11734374 Nigerian_organized_crime, pem_pos=1, final_scores=-0.9371768, final_local_scores=-0.89643675, pem=-0.6912312, similarity=-18.902725]

gt: 22625 Organized_crime, pem_pos=0, final_scores=-1.1068875, final_local_scores=-1.1285965, pem=-0.011989416, similarity=-29.52466]

3: 1521568 Gazprom_Neft

[winner: 1521568 Gazprom_Neft, pem_pos=0, final_scores=-0.92241234, final_local_scores=-1.012044, pem=0.0, similarity=-26.111696]

gt: 1521568 Gazprom_Neft, pem_pos=0, final_scores=-0.92241234, final_local_scores=-1.012044, pem=0.0, similarity=-26.111696]

4: 416486 Business_magnate

[winner: 416486 Business_magnate, pem_pos=0, final_scores=-1.2457671, final_local_scores=-1.2910867, pem=-0.19422136, similarity=-33.382683]

gt: 416486 Business_magnate, pem_pos=0, final_scores=-1.2457671, final_local_scores=-1.2910867, pem=-0.19422136, similarity=-33.382683]

5: 355478 Boris_Berezovsky_(businessman)

[winner: 30237731 Boris_Berezovsky, pem_pos=0,
final_scores=0.09018521, final_local_scores=0.32435316,
pem=0.0, similarity=13.769058]
RECALL MISS: 355478 Boris_Berezovsky_(businessman)]
6: 4207510 Oil
[winner: 540448 Product_(chemistry), pem_pos=0,
final_scores=-0.9435786, final_local_scores=-0.8313806,
pem=0.0, similarity=-20.720343]
RECALL MISS: 4207510 Oil]
[winner: 4207510 Oil, pem_pos=0, final_scores=-1.054277,
final_local_scores=-1.0993664, pem=-0.8430055,
similarity=-24.13318]
gt: 4207510 Oil, pem_pos=0, final_scores=-1.054277,
final_local_scores=-1.0993664, pem=-0.8430055,
similarity=-24.13318]
7: 5987 Coal
[winner: 5987 Coal, pem_pos=0, final_scores=-1.1761483,
final_local_scores=-1.2530141, pem=-0.05381827,
similarity=-33.010056]
gt: 5987 Coal, pem_pos=0, final_scores=-1.1761483,
final_local_scores=-1.2530141, pem=-0.05381827,
similarity=-33.010056]
8: 208413 Coal_mining
[winner: 208413 Coal_mining, pem_pos=0,
final_scores=-1.4347298, final_local_scores=-1.5501652,
pem=-0.040822018, similarity=-41.94831]
gt: 208413 Coal_mining, pem_pos=0,
final_scores=-1.4347298, final_local_scores=-1.5501652,
pem=-0.040822018, similarity=-41.94831]
9: 250879 Contract_killing
[winner: 250879 Contract_killing, pem_pos=0,
final_scores=-1.1377518, final_local_scores=-1.1426508,
pem=0.0, similarity=-30.009264]
gt: 250879 Contract_killing, pem_pos=0,
final_scores=-1.1377518, final_local_scores=-1.1426508,
pem=0.0, similarity=-30.009264]
10: 737234 Mobsters
[winner: 737234 Mobsters, pem_pos=0,
final_scores=-1.1226543, final_local_scores=-1.048612,
pem=0.0, similarity=-27.202961]
gt: 737234 Mobsters, pem_pos=0, final_scores=-1.1226543,
final_local_scores=-1.048612, pem=0.0, similarity=-27.202961]
11: 3890804 Leninsk-Kuznetsky_(city)
[winner: 3890804 Leninsk-Kuznetsky_(city), pem_pos=0,
final_scores=-0.09070468, final_local_scores=-0.065418065,
pem=0.0, similarity=2.1375074]
gt: 3890804 Leninsk-Kuznetsky_(city), pem_pos=0,
final_scores=-0.09070468, final_local_scores=-0.065418065,
pem=0.0, similarity=2.1375074]
12: 101146 Mayor
[winner: 25338759 Mayor_of_Manila, pem_pos=0,
final_scores=-1.0357683, final_local_scores=-0.9801399,

pem=0.0, similarity=-25.159616|
gt: 101146 Mayor, pem_pos=1, final_scores=-1.2840986,
final_local_scores=-1.3200049, pem=-0.98082924,
similarity=-29.967968]
13: 2256795 Tax_evasion
[winner: 2902209 Securities_fraud, pem_pos=1,
final_scores=-1.1163048, final_local_scores=-1.0837157,
pem=-1.0910568, similarity=-22.31719|
RECALL MISS: 2256795 Tax_evasion]
[winner: 215760 Tax_horsepower, pem_pos=1,
final_scores=-1.2491893, final_local_scores=-1.2121,
pem=-0.6929966, similarity=-28.313147|
RECALL MISS: 2256795 Tax_evasion]
14: 209411 Embezzlement
[winner: 209411 Embezzlement, pem_pos=0,
final_scores=-1.231645, final_local_scores=-1.2523224,
pem=0.0, similarity=-33.282085|
gt: 209411 Embezzlement, pem_pos=0,
final_scores=-1.231645, final_local_scores=-1.2523224,
pem=0.0, similarity=-33.282085]
15: 413447 Russian_mafia
[winner: 41328307 The_Mob_(American_rock_band),
pem_pos=0, final_scores=-0.89699656,
final_local_scores=-0.76432866, pem=-0.47425738,
similarity=-16.140293|
RECALL MISS: 413447 Russian_mafia]
16: 3276812 Trial
[winner: 3276812 Trial, pem_pos=0,
final_scores=-1.3883598, final_local_scores=-1.4161484,
pem=-0.32654983, similarity=-36.39515|
gt: 3276812 Trial, pem_pos=0, final_scores=-1.3883598,
final_local_scores=-1.4161484, pem=-0.32654983,
similarity=-36.39515]
1: 17131 Kemerovo
[pem_pos=0, final_scores=0.2946723,
final_local_scores=0.29177368, pem=0.0, similarity=12.796822]
2: 29639 Siberia
[pem_pos=0, final_scores=0.49557757,
final_local_scores=0.5022046, pem=-0.053716753,
similarity=19.36862]
3: 25391 Russia
[pem_pos=0, final_scores=0.8529325,
final_local_scores=0.8376661, pem=-0.20299491,
similarity=30.191257]
4: 2265916 Gennady_Konyakhin
[pem_pos=0, final_scores=0.1472647,
final_local_scores=0.2862903, pem=0.0, similarity=12.633186]
1: 29639 Siberia
[pem_pos=0, final_scores=0.3865173,
final_local_scores=0.3830728, pem=-0.47256553,
similarity=18.091253]
2: 25391 Russia

[pem_pos=1, final_scores=0.8221814,
 final_local_scores=0.80407524, pem=-1.0515022,
 similarity=33.803154]
 3: 25391 Russia
 [pem_pos=1, final_scores=0.829403,
 final_local_scores=0.8119638, pem=-1.0515022,
 similarity=34.038563]
 4: 1521568 Gazprom_Neft
 [pem_pos=0, final_scores=0.41870278,
 final_local_scores=0.47276086, pem=0.0, similarity=18.19784]
 5: 30237731 Boris_Berezovsky
 [pem_pos=0, final_scores=0.09018521,
 final_local_scores=0.32435316, pem=0.0, similarity=13.769058]
 6: 1521568 Gazprom_Neft
 [pem_pos=0, final_scores=0.21915923,
 final_local_scores=0.25478986, pem=0.0, similarity=11.6931505]
 7: 17131 Kemerovo
 [pem_pos=0, final_scores=0.19214775,
 final_local_scores=0.17978114, pem=0.0, similarity=9.454741]
 8: 25391 Russia
 [pem_pos=0, final_scores=0.76776224,
 final_local_scores=0.7446306, pem=-0.20299491,
 similarity=27.414892]
 9: 25391 Russia
 [pem_pos=0, final_scores=0.7959674,
 final_local_scores=0.7754405, pem=-0.20299491,
 similarity=28.334322]

Sample: b'APW19981119_0585_htm'

This `` Dracula '' kills 1 cattle and deer A giant black 2 be
 ar , whom 3 peasants have nicknamed `` Dracula , '' is
 creating anxiety in western [1 1 Romania 1] because of his
 taste for cattle and 4 sheep , an official said Thursday .
 Peasants whose 5 livestock has been devoured by the hungry
 bear have lodged seven complaints in the past six weeks ,
 said Nicolae Onetiu , head of the state hunting and fishing 6
 reserves in western [1 Romania 1] . 7 Dracula was named after
 the [2 2 Transylvania 2] n 8 vampire count of legend because
 of his bloody appetite , said Onetiu . The male bruin weighs
 at least 300 9 kilograms -LRB- 660 pounds -RRB- and has moved
 closer to inhabited areas from forested regions with the
 onset of cold weather , but is apparently not yet ready to 10
 hibernate . Dracula has killed at least six cows and seven
 sheep over the past six weeks . Most bears are 11
 omnivorous , but in [2 Romania 2] , many bears were 12 weaned
 on meat , said Onetiu . [3 3 Nicolae Ceausescu 3] , the 13 co
 mmunist 14 dictator toppled and executed almost a decade
 ago , was an avid hunter , and hunting reserve cubs were fed
 meat to make them grow bigger quicker , leaving generations
 partial to meat . State 15 rangers count 135 bears living in
 the [4 4 Apuseni Mountains 4] in western [3 Romania 3] , all
 of them protected by law . -LRB- str/am/gj -RRB-

1: 25445 Romania
 2: 30853 Transylvania
 3: 49562 Nicolae_Ceaușescu
 4: 1876231 Apuseni_Mountains
 1: 26051975 Cattle
 [winner: 13184204 Baggara_cattle, pem_pos=6,
 final_scores=-1.1458094, final_local_scores=-1.0646262,
 pem=-1.7917595, similarity=-17.93699]
 gt: 26051975 Cattle, pem_pos=0, final_scores=-1.2077835,
 final_local_scores=-1.2402658, pem=-0.012779907,
 similarity=-32.8528]
 2: 4400 Bear
 [winner: 17789992 Bear_in_heraldry, pem_pos=5,
 final_scores=-1.1768633, final_local_scores=-1.0988797,
 pem=-1.9364153, similarity=-18.172522]
 gt: 4400 Bear, pem_pos=0, final_scores=-1.3983715,
 final_local_scores=-1.4599192, pem=-0.2976884,
 similarity=-37.85831]
 3: 68282 Peasant
 [winner: 644162 Autumn_Harvest_Uprising, pem_pos=0,
 final_scores=-0.9757677, final_local_scores=-0.9213674,
 pem=0.0, similarity=-23.405727]
 gt: 68282 Peasant, pem_pos=1, final_scores=-1.2889957,
 final_local_scores=-1.4113872, pem=-0.03302571,
 similarity=-37.8493]
 4: 17158563 Sheep
 [winner: 22787429 Arapawa_Sheep, pem_pos=1,
 final_scores=-0.9730402, final_local_scores=-0.8690254,
 pem=-1.0970988, similarity=-15.877546]
 gt: 17158563 Sheep, pem_pos=0, final_scores=-1.1162065,
 final_local_scores=-1.1406281, pem=0.0,
 similarity=-29.948906]
 5: 25160767 Livestock
 [winner: 25160767 Livestock, pem_pos=0,
 final_scores=-1.3003423, final_local_scores=-1.3394704,
 pem=-0.010359536, similarity=-35.826416]
 gt: 25160767 Livestock, pem_pos=0,
 final_scores=-1.3003423, final_local_scores=-1.3394704,
 pem=-0.010359536, similarity=-35.826416]
 6: 508017 Game_reserve
 [winner: 30035815
 Liga_Nacional_de_Fútbol_de_Honduras_Reserves, pem_pos=8,
 final_scores=-1.6514319, final_local_scores=-1.6308486,
 pem=-2.5304842, similarity=-30.816898]
 RECALL MISS: 508017 Game_reserve]
 7: 7923 Dracula
 [winner: 7923 Dracula, pem_pos=0,
 final_scores=-0.12389059, final_local_scores=-0.07894837,
 pem=-0.48277682, similarity=4.3591514]
 gt: 7923 Dracula, pem_pos=0, final_scores=-0.12389059,
 final_local_scores=-0.07894837, pem=-0.48277682,
 similarity=4.3591514]

8: 32362 Vampire
[winner: 32362 Vampire, pem_pos=0,
final_scores=-0.9760122, final_local_scores=-1.0205188,
pem=-0.137359, similarity=-25.617626]
gt: 32362 Vampire, pem_pos=0, final_scores=-0.9760122,
final_local_scores=-1.0205188, pem=-0.137359,
similarity=-25.617626]

9: 16619 Kilogram
[winner: 16619 Kilogram, pem_pos=0,
final_scores=-0.994752, final_local_scores=-1.0020204,
pem=-0.00055692205, similarity=-25.809542]
gt: 16619 Kilogram, pem_pos=0, final_scores=-0.994752,
final_local_scores=-1.0020204, pem=-0.00055692205,
similarity=-25.809542]

10: 71088 Hibernation
[winner: 3220132 Hibernation_(computing), pem_pos=1,
final_scores=-1.3261125, final_local_scores=-1.2801719,
pem=-1.3628359, similarity=-26.701853]
gt: 71088 Hibernation, pem_pos=0,
final_scores=-1.3908236, final_local_scores=-1.4221694,
pem=-0.35123494, similarity=-36.440594]

11: 23974535 Omnivore
[winner: 23974535 Omnivore, pem_pos=0,
final_scores=-1.2813054, final_local_scores=-1.2712522,
pem=-0.0023532761, similarity=-33.83419]
gt: 23974535 Omnivore, pem_pos=0,
final_scores=-1.2813054, final_local_scores=-1.2712522,
pem=-0.0023532761, similarity=-33.83419]

12: 348349 Weaning
[winner: 348349 Weaning, pem_pos=1,
final_scores=-1.298426, final_local_scores=-1.2627964,
pem=-0.03190216, similarity=-33.421158]
gt: 348349 Weaning, pem_pos=1, final_scores=-1.298426,
final_local_scores=-1.2627964, pem=-0.03190216,
similarity=-33.421158]

13: 9209651 Communism
[winner: 38776970 Communist_Party_of_Cuba, pem_pos=0,
final_scores=-0.8651239, final_local_scores=-0.88538355,
pem=0.0, similarity=-22.331898]
RECALL MISS: 9209651 Communism]

14: 8409 Dictator
[winner: 38776970 Communist_Party_of_Cuba, pem_pos=0,
final_scores=-0.8651239, final_local_scores=-0.88538355,
pem=0.0, similarity=-22.331898]
RECALL MISS: 8409 Dictator]

15: 499846 Park_ranger
[winner: 538185 Ranger_(Middle-earth), pem_pos=10,
final_scores=-1.54216, final_local_scores=-1.4987657,
pem=-3.958439, similarity=-19.109838]
gt: 499846 Park_ranger, pem_pos=15,
final_scores=-2.1334734, final_local_scores=-2.2501678,
pem=-4.025664, similarity=-41.167595]

1: 25445 Romania
 [pem_pos=0, final_scores=0.7271736,
 final_local_scores=0.69422585, pem=-0.21241479,
 similarity=25.96194]
 2: 30853 Transylvania
 [pem_pos=0, final_scores=0.34666103,
 final_local_scores=0.3126179, pem=-0.067388475,
 similarity=13.785324]
 3: 49562 Nicolae_Ceaușescu
 [pem_pos=0, final_scores=0.638334,
 final_local_scores=0.68949515, pem=0.0, similarity=24.665623]
 4: 1876231 Apuseni Mountains
 [pem_pos=0, final_scores=0.18478678,
 final_local_scores=0.2276339, pem=0.0, similarity=10.882762]
 1: 25445 Romania
 [pem_pos=0, final_scores=0.69901866,
 final_local_scores=0.66347075, pem=-0.21241479,
 similarity=25.044147]
 2: 25445 Romania
 [pem_pos=0, final_scores=0.64123386,
 final_local_scores=0.60034966, pem=-0.21241479,
 similarity=23.160488]
 3: 25445 Romania
 [pem_pos=0, final_scores=0.67591673,
 final_local_scores=0.63823545, pem=-0.21241479,
 similarity=24.291077]

Sample: b'APW19980930_0522_htm'

[1 Iraqi 1] minister arrives in [2 Iran 2] to attend 1 trade fair [1 1 Iraq 1] 's trade minister arrived Wednesday to attend an international trade fair and hold discussions with his [2 2 Iranian 2] counterpart , the [3 Islamic Republic News Agency 3] reported . Mohammed Mehdi Saleh and his seven-member 2 delegation traveled by land and crossed into [4 Iran 4] at the Khosrawi border point in the western [5 3 Kermanshah 5] province , the official news agency said . It quoted [6 Saleh 6] as saying that he would discuss with [7 Iran 7] 's Trade Minister 4 Mohammad Shariat-Madari ways of expanding 5 commercial ties between the two countries . The fair is scheduled from Oct. 1 to 11 . This is the second consecutive year that [8 Iraq 8] is attending the 6 Tehran International Trade Fair , and points at the need for the two neighbors to do business despite their traditional animosity . [9 Iran 9] and [10 Iraq 10] fought an 7 eight-year war which ended with a [3 3 U.N. 3] - brokered cease-fire . [11 Iran 11] claims [12 Iraq 12] still holds 4,000 8 prisoners of war while [13 Iraq 1 3] says [14 Iran 14] is holding about 20,000 prisoners . More than 1 million people were killed or wounded on both sides during the 1980-88 war . [15 Tehran 15] accuses [16 9 Baghdad 16] of helping [17 Iranian 17] opposition groups and the two sides accuse each other of allowing infiltrators to sneak across their borders for 10 acts of sabotage . In December , [

18 Iraqi 18] Vice President [4 4 Taha Yassin Ramadan 4]
became the highest-ranking [19 Iraqi 19] to visit [20 Tehran 2
0] since the 5 1991 [5 Gulf War 5] when he attended the [6 6 0
rganization of the Islamic Conference 6] meeting . ti/vj

1: 7515928 Iraq

2: 14653 Iran

3: 31769 United_Nations

4: 203183 Taha_Yassin_Ramadan

5: 182000 Gulf_War

6: 47725 Organisation_of_Islamic_Cooperation

1: 1129368 Trade_fair

[winner: 1528883 Aircraft_fairing, pem_pos=4,
final_scores=-1.1097604, final_local_scores=-1.0112187,
pem=-1.3849394, similarity=-18.555553]

RECALL MISS: 1129368 Trade_fair]

[winner: 1090151 Border_trade, pem_pos=5,
final_scores=-1.4284396, final_local_scores=-1.469475,
pem=-2.0768182, similarity=-28.46829]

RECALL MISS: 1129368 Trade_fair]

2: 218837 Delegation

[winner: 13540279 Delegations_of_Tunisia, pem_pos=2,
final_scores=-1.1569022, final_local_scores=-1.1534872,
pem=-1.7203006, similarity=-20.977383]

gt: 218837 Delegation, pem_pos=0,
final_scores=-1.4516425, final_local_scores=-1.4999847,
pem=-0.6624647, similarity=-37.07024]

3: 448078 Kermanshah_Province

[winner: 1489244 Kermanshah, pem_pos=0,
final_scores=0.3277483, final_local_scores=0.3109593,
pem=0.0, similarity=13.369359]

gt: 448078 Kermanshah_Province, pem_pos=1,
final_scores=-0.0987563, final_local_scores=-0.060348988,
pem=-0.5025201, similarity=5.021561]

[winner: 673434 Provinces_of_Burkina_Faso, pem_pos=3,
final_scores=-1.5517344, final_local_scores=-1.5204613,
pem=-2.634314, similarity=-26.958073]

RECALL MISS: 448078 Kermanshah_Province]

4: 38597414 Mohammad_Shariatmadari

[winner: 18934 Muhammad, pem_pos=0,
final_scores=-1.3011622, final_local_scores=-1.3986998,
pem=-0.17946821, similarity=-36.6743]

RECALL MISS: 38597414 Mohammad_Shariatmadari]

5: 39208 Commerce

[winner: 1621419 Ensign-Bickford_Company, pem_pos=11,
final_scores=-1.1387631, final_local_scores=-1.0081182,
pem=-2.4842746, similarity=-12.484674]

gt: 39208 Commerce, pem_pos=0, final_scores=-1.7579986,
final_local_scores=-1.8629112, pem=-0.6983233,
similarity=-47.705685]

6: 57654 Tehran

[winner: 57654 Tehran, pem_pos=0,
final_scores=-0.80000657, final_local_scores=-0.9348347,

pem=0.0, similarity=-23.80762|
gt: 57654 Tehran, pem_pos=0, final_scores=-0.80000657,
final_local_scores=-0.9348347, pem=0.0, similarity=-23.80762]
7: 14889 Iran-Iraq_War
[winner: 14889 Iran-Iraq_War, pem_pos=0,
final_scores=-1.1351165, final_local_scores=-1.2889061,
pem=0.0, similarity=-34.373817|
gt: 14889 Iran-Iraq_War, pem_pos=0,
final_scores=-1.1351165, final_local_scores=-1.2889061,
pem=0.0, similarity=-34.373817]
8: 25008 Prisoner_of_war
[winner: 14244852 Nazi_crimes_against_Soviet_POWs,
pem_pos=1, final_scores=-1.2532918,
final_local_scores=-1.2570353, pem=-0.6829134,
similarity=-29.708939|
gt: 25008 Prisoner_of_war, pem_pos=0,
final_scores=-1.3934883, final_local_scores=-1.49664,
pem=0.0, similarity=-40.573006]
9: 4492 Baghdad
[winner: 7515928 Iraq, pem_pos=8, final_scores=0.1461965,
final_local_scores=0.06273168, pem=-2.5644467,
similarity=19.907612|
gt: 4492 Baghdad, pem_pos=0, final_scores=0.11130707,
final_local_scores=0.05196072, pem=0.0, similarity=5.6403246]
10: 29462 Sabotage
[winner: 29462 Sabotage, pem_pos=0,
final_scores=-1.1660782, final_local_scores=-1.2113631,
pem=-0.019294495, similarity=-31.954851|
gt: 29462 Sabotage, pem_pos=0, final_scores=-1.1660782,
final_local_scores=-1.2113631, pem=-0.019294495,
similarity=-31.954851]
[winner: 2084 Acts_of_the_Apostles, pem_pos=0,
final_scores=-1.4495485, final_local_scores=-1.4720964,
pem=0.0, similarity=-39.84058|
RECALL MISS: 29462 Sabotage]
1: 7515928 Iraq
[pem_pos=0, final_scores=0.6028022,
final_local_scores=0.559971, pem=-0.25379938,
similarity=22.180565]
2: 14653 Iran
[pem_pos=0, final_scores=0.5169794,
final_local_scores=0.49625707, pem=-0.57066226,
similarity=22.002363]
3: 31769 United_Nations
[pem_pos=0, final_scores=0.47352225,
final_local_scores=0.5146556, pem=0.0, similarity=19.448063]
4: 203183 Taha_Yassin_Ramadan
[pem_pos=0, final_scores=0.23174028,
final_local_scores=0.31837708, pem=0.0, similarity=13.59072]
5: 182000 Gulf_War
[pem_pos=0, final_scores=0.3362664,
final_local_scores=0.34793052, pem=0.0, similarity=14.472653]

6: 47725 Organisation_of_Islamic_Cooperation
[pem_pos=0, final_scores=0.27401572,
final_local_scores=0.31374025, pem=0.0, similarity=13.452348]
1: 7515928 Iraq
[pem_pos=0, final_scores=0.55098015,
final_local_scores=0.5033632, pem=-0.48469657,
similarity=21.74693]
2: 14653 Iran
[pem_pos=0, final_scores=0.48283738,
final_local_scores=0.45896214, pem=-0.13856095,
similarity=18.539576]
3: 841314 Islamic_Republic_News_Agency
[pem_pos=0, final_scores=0.45053995,
final_local_scores=0.5042358, pem=0.0, similarity=19.137115]
4: 14653 Iran
[pem_pos=0, final_scores=0.52000135,
final_local_scores=0.49955815, pem=-0.13856095,
similarity=19.751041]
5: 1489244 Kermanshah
[pem_pos=0, final_scores=0.3277483,
final_local_scores=0.3109593, pem=0.0, similarity=13.369359]
6: 215609 Ali_Abdullah_Saleh
[pem_pos=1, final_scores=0.10530351,
final_local_scores=0.12655714, pem=-0.95323664,
similarity=13.05028]
7: 14653 Iran
[pem_pos=0, final_scores=0.64911354,
final_local_scores=0.6405936, pem=-0.13856095,
similarity=23.95982]
8: 7515928 Iraq
[pem_pos=0, final_scores=0.6035412,
final_local_scores=0.56077814, pem=-0.25379938,
similarity=22.204653]
9: 14653 Iran
[pem_pos=0, final_scores=0.61448807,
final_local_scores=0.6027705, pem=-0.13856095,
similarity=22.831104]
10: 7515928 Iraq
[pem_pos=0, final_scores=0.61095303,
final_local_scores=0.5688744, pem=-0.25379938,
similarity=22.44626]
11: 14653 Iran
[pem_pos=0, final_scores=0.56702614,
final_local_scores=0.55092555, pem=-0.13856095,
similarity=21.283947]
12: 7515928 Iraq
[pem_pos=0, final_scores=0.61674684,
final_local_scores=0.5752033, pem=-0.25379938,
similarity=22.635128]
13: 7515928 Iraq
[pem_pos=0, final_scores=0.5894145,
final_local_scores=0.54534686, pem=-0.25379938,

```

similarity=21.744152]
14: 14653 Iran
    [pem_pos=0, final_scores=0.7021714,
final_local_scores=0.69855124, pem=-0.13856095,
similarity=25.689392]
15: 14653 Iran
    [pem_pos=5, final_scores=0.15906058,
final_local_scores=0.10396114, pem=-2.6998966,
similarity=21.874578]
16: 7515928 Iraq
    [pem_pos=8, final_scores=0.1461965,
final_local_scores=0.06273168, pem=-2.5644467,
similarity=19.907612]
17: 14653 Iran
    [pem_pos=0, final_scores=0.5441535,
final_local_scores=0.5259407, pem=-0.57066226,
similarity=22.888182]
18: 7515928 Iraq
    [pem_pos=0, final_scores=0.5775928,
final_local_scores=0.5324335, pem=-0.48469657,
similarity=22.614447]
19: 7515928 Iraq
    [pem_pos=0, final_scores=0.51503134,
final_local_scores=0.46409458, pem=-0.48469657,
similarity=20.575077]
20: 57654 Tehran
    [pem_pos=0, final_scores=0.108706996,
final_local_scores=0.06181115, pem=0.0, similarity=5.934281]

```

Sample: b'APW19990203_0315_htm'

Museum To Return 3 Stolen Works [1 1 LOS ANGELES 1] -LRB- AP -RRB- -- 2 The [2 J. Paul Getty Museum 2] is returning three ancient 1 works of art to [3 3 Italy 3] after determining they were stolen . One 2 sculpture sat in its 3 collection for 17 years . The objects will be flown to [1 Italy 1] this week , along with pieces that were loaned to the museum for an 4 antiquities exhibit , according to a statement issued by the [2 Getty 2] on Monday . The stolen items are a terra cotta drinking cup from 480 5 B.C. , found to have been illegally 6 excavated from an 7 archaeological site ; a 8 torso of the god 9 Mithras from the 10 2nd century , found to have been stolen from a private [3 Italian 3] collection ; and the sculpted head of a victorious athlete , a 2nd century 11 Roman copy of an earlier [4 12 Greek 4] work , which was apparently taken from the storeroom of a scientific excavation . For security reasons , the museum did not say how much the pieces are worth . Although nobody claimed the works , Getty decided to return them based on information supplied by experts in the field and in consultation with [5 Italian 5] officials . International concern has been growing about rampant 13 thefts of priceless art from collections and the 14 looting of archaeological sites . The billion-dollar

museum , located on a hill in the 15 Brentwood area , opened in 1997 .

1: 18110 Los_Angeles

2: 1355493 J._Paul_Getty_Museum

3: 14532 Italy

1: 2583747 Work_of_art

[winner: 752 Art, pem_pos=0, final_scores=-1.2167407, final_local_scores=-1.3385453, pem=-0.22415501, similarity=-34.63616]

gt: 2583747 Work_of_art, pem_pos=2, final_scores=-1.6067728, final_local_scores=-1.7383683, pem=-2.7870064, similarity=-32.63049]

[winner: 16892711 Charles_Ward_Engineering_Works, pem_pos=8, final_scores=-1.2641089, final_local_scores=-1.1970947, pem=-3.0445223, similarity=-15.077393]

RECALL MISS: 2583747 Work_of_art]

2: 26714 Sculpture

[winner: 1934053 Sculpture_(mollusc), pem_pos=1, final_scores=-1.0428581, final_local_scores=-0.9592969, pem=-1.1696436, similarity=-18.176914]

gt: 26714 Sculpture, pem_pos=0, final_scores=-1.1526705, final_local_scores=-1.2846532, pem=-0.17347881, similarity=-33.303493]

3: 3218952 Collection_(artwork)

[winner: 25004018

United_States_Senate_Vice_Presidential_Bust_Collection, pem_pos=0, final_scores=-0.8460935, final_local_scores=-0.7802044, pem=0.0, similarity=-19.193144]

RECALL MISS: 3218952 Collection_(artwork)]

4: 1431051 Antiquities

[winner: 1431051 Antiquities, pem_pos=0, final_scores=-1.3055173, final_local_scores=-1.4328258, pem=-0.12439071, similarity=-37.99221]

gt: 1431051 Antiquities, pem_pos=0, final_scores=-1.3055173, final_local_scores=-1.4328258, pem=-0.12439071, similarity=-37.99221]

5: 1400 Anno_Domini

[winner: 3392 British_Columbia, pem_pos=2, final_scores=-0.2265239, final_local_scores=-0.25067234, pem=-0.84342366, similarity=1.1958199]

gt: 1400 Anno_Domini, pem_pos=1, final_scores=-0.33217388, final_local_scores=-0.34273973, pem=-0.7285973, similarity=-2.1760993]

6: 60697 Excavation_(archaeology)

[winner: 60697 Excavation_(archaeology), pem_pos=0, final_scores=-1.2237952, final_local_scores=-1.3192, pem=0.0, similarity=-35.277847]

gt: 60697 Excavation_(archaeology), pem_pos=0, final_scores=-1.2237952, final_local_scores=-1.3192, pem=0.0, similarity=-35.277847]

7: 147625 Archaeological_site
 [winner: 18083113 Mitchell_Site, pem_pos=0,
 final_scores=-0.908775, final_local_scores=-0.8090072,
 pem=0.0, similarity=-20.052677|
 RECALL MISS: 147625 Archaeological_site]

8: 674207 Trunk_(anatomy)
 [winner: 674207 Trunk_(anatomy), pem_pos=0,
 final_scores=-1.4320619, final_local_scores=-1.4455731,
 pem=0.0, similarity=-39.04907|
 gt: 674207 Trunk_(anatomy), pem_pos=0,
 final_scores=-1.4320619, final_local_scores=-1.4455731,
 pem=0.0, similarity=-39.04907]

9: 20826 Mithraic_mysteries
 [winner: 20826 Mithraic_mysteries, pem_pos=0,
 final_scores=-0.029977426, final_local_scores=0.036713883,
 pem=0.0, similarity=5.1853285|
 gt: 20826 Mithraic_mysteries, pem_pos=0,
 final_scores=-0.029977426, final_local_scores=0.036713883,
 pem=0.0, similarity=5.1853285]

10: 34744 2nd_century
 [winner: 34744 2nd_century, pem_pos=0,
 final_scores=-0.66167694, final_local_scores=-0.593168,
 pem=0.0, similarity=-13.611605|
 gt: 34744 2nd_century, pem_pos=0,
 final_scores=-0.66167694, final_local_scores=-0.593168,
 pem=0.0, similarity=-13.611605]

11: 521555 Ancient_Rome
 [winner: 521555 Ancient_Rome, pem_pos=0,
 final_scores=-0.04496233, final_local_scores=-0.060139313,
 pem=-1.0249887, similarity=7.8690834|
 gt: 521555 Ancient_Rome, pem_pos=0,
 final_scores=-0.04496233, final_local_scores=-0.060139313,
 pem=-1.0249887, similarity=7.8690834]

12: 66540 Ancient_Greece
 [winner: 12108 Greece, pem_pos=1,
 final_scores=0.35153264, final_local_scores=0.3451447,
 pem=-2.3247876, similarity=27.032076|
 gt: 66540 Ancient_Greece, pem_pos=4,
 final_scores=-0.31645367, final_local_scores=-0.34128082,
 pem=-2.8341312, similarity=9.317658]

13: 31287 Theft
 [winner: 31287 Theft, pem_pos=0, final_scores=-1.4179444,
 final_local_scores=-1.4775677, pem=0.0,
 similarity=-40.003853|
 gt: 31287 Theft, pem_pos=0, final_scores=-1.4179444,
 final_local_scores=-1.4775677, pem=0.0,
 similarity=-40.003853]

14: 209811 Looting
 [winner: 209811 Looting, pem_pos=0,
 final_scores=-1.3720905, final_local_scores=-1.452007,
 pem=-0.0523326, similarity=-38.956482|
 gt: 209811 Looting, pem_pos=0, final_scores=-1.3720905,

```

final_local_scores=-1.452007, pem=-0.0523326,
similarity=-38.956482 ]
15: 345873 Brentwood, Los Angeles
    [winner: 345873 Brentwood, Los Angeles, pem_pos=0,
final_scores=-0.036901712, final_local_scores=-0.009733319,
pem=-1.7519233, similarity=13.326483]
    gt: 345873 Brentwood, Los Angeles, pem_pos=0,
final_scores=-0.036901712, final_local_scores=-0.009733319,
pem=-1.7519233, similarity=13.326483 ]
1: 18110 Los Angeles
    [pem_pos=0, final_scores=0.1361611,
final_local_scores=0.12925664, pem=-0.100652285,
similarity=8.494351]
2: 1355493 J. Paul Getty Museum
    [pem_pos=0, final_scores=0.23616312,
final_local_scores=0.2918292, pem=0.0, similarity=12.798478]
3: 14532 Italy
    [pem_pos=0, final_scores=0.70743704,
final_local_scores=0.68843, pem=-0.20612453,
similarity=25.754776]
1: 14532 Italy
    [pem_pos=0, final_scores=0.6733685,
final_local_scores=0.6512153, pem=-0.20612453,
similarity=24.644215]
2: 174527 J. Paul Getty
    [pem_pos=0, final_scores=0.0920098,
final_local_scores=0.14056376, pem=0.0, similarity=8.284416]
3: 14532 Italy
    [pem_pos=0, final_scores=0.71699196,
final_local_scores=0.6988673, pem=-0.75948656,
similarity=29.075514]
4: 12108 Greece
    [pem_pos=1, final_scores=0.35153264,
final_local_scores=0.3451447, pem=-2.3247876,
similarity=27.032076]
5: 14532 Italy
    [pem_pos=0, final_scores=0.7402215,
final_local_scores=0.7242421, pem=-0.75948656,
similarity=29.832747]

```

Sample: b'APW19980930_0284_htm'

[1 Impressionist 1] paintings recovered , 1 curator central suspect Police have recovered two valuable [1 1 Impressionist 1] paintings stolen 10 days ago in what appears to have been a staged 2 heist and arrested the museum curator for alleged armed robbery , police sources said Wednesday . Two other people , described as small-time thugs , also were arrested in connection with the Sept. 21 theft of a [2 2 Monet 2] and a [2 Sisley 2] at the Fine Arts Museum of 3 Nice , according to the sources in the judicial police . They spoke on customary condition of anonymity . The art works , [3 Claude Monet 3] 's 1897 painting `` The Cliffs of Dieppe '' and [3 3

Alfred Sisley 3] 's 1890 work `` The Alley of the Poplars ,
' ' were found on a sail boat anchored in the small [4 4 Rivie
ra 4] port of [5 5 Saint Laurent du Var 5] , police said .
The police sources said that curator Jean Forneris is
suspected of having ordered the theft . The paintings were
stolen by two masked men who took Forneris from his home ,
drove him to the museum , then bound and gagged him , the
caretaker and another employee and shut them in the museum
library . The museum alarm had been turned off , allegedly
because the caretaker was on duty . The two men sped off in
the curator 's car with the paintings . Forneris later
expressed his `` shock ' ' at the thefts . -LRB- eg-parf -RRB-

1: 15169 Impressionism

2: 6548 Claude_Monet

3: 291939 Alfred_Sisley

4: 266388 French_Riviera

5: 1077214 Saint-Laurent-du-Var

1: 391167 Curator

[winner: 391167 Curator, pem_pos=0,
final_scores=-1.5207825, final_local_scores=-1.6446962,
pem=-0.020542735, similarity=-44.87959|

gt: 391167 Curator, pem_pos=0, final_scores=-1.5207825,
final_local_scores=-1.6446962, pem=-0.020542735,
similarity=-44.87959]

2: 54849 Robbery

[winner: 13419935 Heist,_Germany, pem_pos=2,
final_scores=-0.6975276, final_local_scores=-0.49387446,
pem=-1.7161669, similarity=-1.3157043|

RECALL MISS: 54849 Robbery]

3: 47088 Nice

[winner: 47088 Nice, pem_pos=0, final_scores=-0.7261788,
final_local_scores=-0.81494, pem=-0.2595403,
similarity=-18.8183|

gt: 47088 Nice, pem_pos=0, final_scores=-0.7261788,
final_local_scores=-0.81494, pem=-0.2595403,
similarity=-18.8183]

1: 15169 Impressionism

[pem_pos=0, final_scores=0.2582718,
final_local_scores=0.19868839, pem=0.0, similarity=10.0189705]

2: 6548 Claude_Monet

[pem_pos=0, final_scores=0.28368086,
final_local_scores=0.2157065, pem=0.0, similarity=10.526825]

3: 291939 Alfred_Sisley

[pem_pos=0, final_scores=0.3477491,
final_local_scores=0.30170602, pem=0.0, similarity=13.093223]

4: 266388 French_Riviera

[pem_pos=3, final_scores=0.1309825,
final_local_scores=0.11789793, pem=-1.8379819,
similarity=17.603254]

5: 1077214 Saint-Laurent-du-Var

[pem_pos=0, final_scores=0.18079333,
final_local_scores=0.27645186, pem=0.0, similarity=12.339588]

1: 15169 Impressionism
 [pem_pos=0, final_scores=0.1917953,
 final_local_scores=0.12607294, pem=0.0, similarity=7.851981]
 2: 291939 Alfred_Sisley
 [pem_pos=0, final_scores=0.21278445,
 final_local_scores=0.15427765, pem=0.0, similarity=8.6936655]
 3: 6548 Claude_Monet
 [pem_pos=0, final_scores=0.38462806,
 final_local_scores=0.32597595, pem=0.0, similarity=13.817486]

Sample: b'APW19981215_1083_htm'

[1 German 1] zoo hopes court will allow it to accept baby elephants The 1 Dresden 2 Zoo went to court Tuesday to challenge 1 Environment Minister [1 Juergen Trittin 1] 's order forbidding the import of four baby 3 elephants taken from the wild in [2 2 Africa 2] and abused . Zoo director Hubert Luecker said the complaint was filed with an administrative court in [3 3 Cologne 3] . The elephants , which were to be sent this month to the [2 Dresden 2] and [4 4 Erfurt 4] zoos in eastern [5 5 Germany 5] , were part of a group of 30 young elephants captured in August from herds in [6 6 Botswana 6] and taken to a [7 7 South African 7] farm . A [3 South African 3] judge ruled early this month that the animals had been abused and deprived of water at the farm . He granted temporary custody to the National Society for the Prevention of Cruelty to Animals . But while the judge approved the export of four baby elephants to [4 Germany 4] and three to [8 8 Switzerland 8] , [5 Trittin 5] last week said [6 Germany 6] would protest the abuse of animals by not accepting them in its zoos . Others argue the elephants may face a worse fate if they do n't come to [7 Germany 7] . Norbert Otto , a federal lawmaker from [8 Erfurt 8] , demanded in 4 parliament Tuesday that [9 Trittin 9] take steps to prevent the animals from being 5 put to sleep in [10 South Africa 10] . In addition , the cultural minister of [9 9 Saxony 9] , of which [11 Dresden 11] is the capital , asked the state 's 740,000 school children to write letters of protest urging Tritten to allow the elephants into [12 Germany 12] . -LRB- aet-pfg -RRB-

1: 216894 Jürgen_Trittin
 2: 5334607 Africa
 3: 6187 Cologne
 4: 9481 Erfurt
 5: 11867 Germany
 6: 3464 Botswana
 7: 17416221 South_Africa
 8: 26748 Switzerland
 9: 28395 Saxony
 1: 37410 Dresden

[winner: 27238505 Dresden_Zoo, pem_pos=0,
 final_scores=-0.20794061, final_local_scores=-0.13346055,
 pem=0.0, similarity=0.106984064]

RECALL MISS: 37410 Dresden]
2: 9096372 Zoo
[winner: 27238505 Dresden_Zoo, pem_pos=0,
final_scores=-0.20794061, final_local_scores=-0.13346055,
pem=0.0, similarity=0.106984064]
RECALL MISS: 9096372 Zoo]
3: 9279 Elephant
[winner: 4180483 Temple_elephant, pem_pos=8,
final_scores=-1.0542272, final_local_scores=-0.9647617,
pem=-2.1958551, similarity=-12.759297]
gt: 9279 Elephant, pem_pos=0, final_scores=-1.2030469,
final_local_scores=-1.2404504, pem=-0.15377009,
similarity=-32.091576]
4: 3768 Bundestag
[winner: 24406 Parliament, pem_pos=1,
final_scores=-1.695456, final_local_scores=-1.8191772,
pem=-1.8263724, similarity=-40.266064]
RECALL MISS: 3768 Bundestag]
5: 631308 Animal_euthanasia
[winner: 8430643 Sleep-deprived_driving, pem_pos=0,
final_scores=-1.1304178, final_local_scores=-1.0750743,
pem=0.0, similarity=-27.99265]
RECALL MISS: 631308 Animal_euthanasia]
[winner: 207383 Put_option, pem_pos=0,
final_scores=-1.2459731, final_local_scores=-1.2339725,
pem=0.0, similarity=-32.73449]
RECALL MISS: 631308 Animal_euthanasia]
1: 216894 Jürgen_Trittin
[pem_pos=0, final_scores=0.58627737,
final_local_scores=0.68964773, pem=0.0, similarity=24.670177]
2: 5334607 Africa
[pem_pos=0, final_scores=0.55997545,
final_local_scores=0.60570055, pem=-0.13197383,
similarity=22.88272]
3: 6187 Cologne
[pem_pos=0, final_scores=0.50238824,
final_local_scores=0.48162568, pem=-0.08239529,
similarity=18.910461]
4: 9481 Erfurt
[pem_pos=0, final_scores=0.39782745,
final_local_scores=0.3693192, pem=0.0, similarity=15.110935]
5: 11867 Germany
[pem_pos=0, final_scores=0.7474707,
final_local_scores=0.70760196, pem=-0.15529114,
similarity=26.050465]
6: 3464 Botswana
[pem_pos=0, final_scores=0.4680189,
final_local_scores=0.5121551, pem=-0.37041438,
similarity=21.387815]
7: 17416221 South_Africa
[pem_pos=0, final_scores=0.39423102,
final_local_scores=0.39871722, pem=-0.1389451,

similarity=16.743835]
8: 26748 Switzerland
[pem_pos=0, final_scores=0.6125576,
final_local_scores=0.5926473, pem=-0.18598008,
similarity=22.786879]
9: 28395 Saxony
[pem_pos=0, final_scores=0.36503375,
final_local_scores=0.33351064, pem=-0.14660512,
similarity=14.839597]
1: 11867 Germany
[pem_pos=1, final_scores=0.76907235,
final_local_scores=0.7311986, pem=-1.0465823,
similarity=31.601618]
2: 37410 Dresden
[pem_pos=0, final_scores=0.5012544,
final_local_scores=0.47547597, pem=-0.2008269,
similarity=19.370993]
3: 17416221 South_Africa
[pem_pos=0, final_scores=0.44255686,
final_local_scores=0.45150584, pem=-0.1389451,
similarity=18.319153]
4: 11867 Germany
[pem_pos=0, final_scores=0.8112722,
final_local_scores=0.7772954, pem=-0.15529114,
similarity=28.130257]
5: 216894 Jürgen_Trittin
[pem_pos=0, final_scores=0.06543569,
final_local_scores=0.12070739, pem=0.0, similarity=7.691862]
6: 11867 Germany
[pem_pos=0, final_scores=0.7015499,
final_local_scores=0.6574405, pem=-0.15529114,
similarity=24.553547]
7: 11867 Germany
[pem_pos=0, final_scores=0.86397463,
final_local_scores=0.8348649, pem=-0.15529114,
similarity=29.848244]
8: 9481 Erfurt
[pem_pos=0, final_scores=0.40167874,
final_local_scores=0.37352616, pem=0.0, similarity=15.236479]
9: 216894 Jürgen_Trittin
[pem_pos=0, final_scores=0.15582506,
final_local_scores=0.21944404, pem=0.0, similarity=10.638361]
10: 17416221 South_Africa
[pem_pos=0, final_scores=0.5072184,
final_local_scores=0.5221387, pem=-0.26251382,
similarity=21.098965]
11: 37410 Dresden
[pem_pos=0, final_scores=0.47041923,
final_local_scores=0.44179326, pem=-0.2008269,
similarity=18.365833]
12: 11867 Germany
[pem_pos=0, final_scores=0.81726354,

final_local_scores=0.78384006, pem=-0.15529114,
similarity=28.325562]

Sample: b'APW19980811_0512_htm'

Police identify substance in 1 tea poisoning case The poison that sickened 10 [1 Japanese 1] company employees after they drank tea brewed from the same water pot was a 2 toxic 3 sodium compound that affects the 4 central nervous system , 5 police said Tuesday . Police still had no clues to who might have put the chemical in the pot , which had been filled with water between Friday night and Monday , they added . The 10 workers at Xyence Co. , a 6 wood preservative company in [1 1 Niigata 1] , 256 kilometers -LRB- 159 miles -RRB- northwest of [2 2 Tokyo 2] , were hospitalized Monday after suffering 7 nausea and 8 numbness during their morning tea break . The case gave [2 Japan 2] another shock two weeks after 9 curry laced with 10 cyanide and 11 arsenic killed four people at a festival in southwestern [3 Japan 3] . It also raised fears of 12 copycat attacks . Lab tests on the tea revealed traces of 13 sodium azide , a chemical used in many industrial products , including 14 air-bag mechanisms , said police spokesman Masashi Toma . The victims of the poisoning were in stable condition , but three suffered mild 15 liver disorders Tuesday , said Yasuo Hirose , an official at the 16 Niigata City Hospital . The 17 March 1995 18 nerve gas attack on [4 Tokyo 4] subways carried out by the [3 3 Aum Shinri Kyo 3] religious cult has made [5 Japan 5] particularly wary of poisoning incidents . The subway gassing killed 12 people . -LRB- js -RRB-

1: 6793142 Niigata,_Niigata

2: 30057 Tokyo

3: 34267361 Aum_Shinrikyo

1: 29969 Tea

[winner: 2500459 Cannabis_tea, pem_pos=0,
final_scores=-0.9812613, final_local_scores=-0.90134776,
pem=0.0, similarity=-22.808302]

RECALL MISS: 29969 Tea]

2: 170567 Toxicity

[winner: 170567 Toxicity, pem_pos=0,
final_scores=-1.279407, final_local_scores=-1.2729487,
pem=-0.20287444, similarity=-32.794357]

gt: 170567 Toxicity, pem_pos=0, final_scores=-1.279407,
final_local_scores=-1.2729487, pem=-0.20287444,
similarity=-32.794357]

3: 26826 Sodium

[winner: 5064081 Sodium_iodate, pem_pos=3,
final_scores=-0.91065687, final_local_scores=-0.781721,
pem=-1.3859764, similarity=-11.701245]

gt: 26826 Sodium, pem_pos=0, final_scores=-1.1630205,
final_local_scores=-1.1268747, pem=-0.07336842,
similarity=-29.139484]

[winner: 1558261 Inorganic_compounds_by_element,

pem_pos=7, final_scores=-0.9322818,
final_local_scores=-0.7730424, pem=-2.0794415,
similarity=-7.6710877|
RECALL MISS: 26826 Sodium]
4: 7251 Central_nervous_system
[winner: 7251 Central_nervous_system, pem_pos=0,
final_scores=-0.99099135, final_local_scores=-0.94359577,
pem=-0.0026450537, similarity=-24.054682|
gt: 7251 Central_nervous_system, pem_pos=0,
final_scores=-0.99099135, final_local_scores=-0.94359577,
pem=-0.0026450537, similarity=-24.054682]
5: 23627 Police
[winner: 2158227 Milicja_Obywatelska, pem_pos=19,
final_scores=-1.5978982, final_local_scores=-1.594301,
pem=-4.1661835, similarity=-20.831053|
gt: 23627 Police, pem_pos=0, final_scores=-1.6061314,
final_local_scores=-1.6956089, pem=-0.116511196,
similarity=-45.877033]
6: 481484 Wood_preservation
[winner: 4782903 Dermacentor_variabilis, pem_pos=5,
final_scores=-0.99106866, final_local_scores=-0.8358373,
pem=-1.609438, similarity=-12.1009655|
RECALL MISS: 481484 Wood_preservation]
[winner: 24862 Preservative, pem_pos=0,
final_scores=-1.1309037, final_local_scores=-1.0998871,
pem=-0.12771058, similarity=-28.038605|
gt: 481484 Wood_preservation, pem_pos=3,
final_scores=-1.7314061, final_local_scores=-1.7464187,
pem=-4.02299, similarity=-26.149256]
7: 18947703 Nausea
[winner: 18947703 Nausea, pem_pos=0,
final_scores=-1.0019968, final_local_scores=-0.9344321,
pem=-0.008996061, similarity=-23.746683|
gt: 18947703 Nausea, pem_pos=0, final_scores=-1.0019968,
final_local_scores=-0.9344321, pem=-0.008996061,
similarity=-23.746683]
8: 398940 Paresthesia
[winner: 398940 Paresthesia, pem_pos=0,
final_scores=-0.8530516, final_local_scores=-0.7537647,
pem=-0.06883712, similarity=-18.029783|
gt: 398940 Paresthesia, pem_pos=0,
final_scores=-0.8530516, final_local_scores=-0.7537647,
pem=-0.06883712, similarity=-18.029783]
9: 6597 Curry
[winner: 18376305 Curry_Island, pem_pos=1,
final_scores=-1.0015352, final_local_scores=-0.8791279,
pem=-2.2986336, similarity=-9.644891|
gt: 6597 Curry, pem_pos=0, final_scores=-1.4090346,
final_local_scores=-1.4250958, pem=-0.85919905,
similarity=-33.76553]
10: 5910 Cyanide
[winner: 5910 Cyanide, pem_pos=0,

final_scores=-1.1408918, final_local_scores=-1.1041229,
pem=-0.08839158, similarity=-28.378828|
gt: 5910 Cyanide, pem_pos=0, final_scores=-1.1408918,
final_local_scores=-1.1041229, pem=-0.08839158,
similarity=-28.378828]
11: 897 Arsenic
[winner: 5361990 Arsenic_contamination_of_groundwater,
pem_pos=0, final_scores=-1.1556015,
final_local_scores=-1.1415377, pem=0.0,
similarity=-29.976046|
gt: 897 Arsenic, pem_pos=1, final_scores=-1.2232916,
final_local_scores=-1.2187376, pem=-0.038114544,
similarity=-32.072575]
12: 10827311 Copycat_crime
[winner: 3154773 Copycat_(film), pem_pos=0,
final_scores=-1.0695115, final_local_scores=-0.9991627,
pem=-0.8186768, similarity=-21.275202|
gt: 10827311 Copycat_crime, pem_pos=8,
final_scores=-1.7103086, final_local_scores=-1.7428162,
pem=-3.325882, similarity=-29.832739]
13: 1172932 Sodium_azide
[winner: 1172932 Sodium_azide, pem_pos=0,
final_scores=-0.6631984, final_local_scores=-0.5359001,
pem=0.0, similarity=-11.9026165|
gt: 1172932 Sodium_azide, pem_pos=0,
final_scores=-0.6631984, final_local_scores=-0.5359001,
pem=0.0, similarity=-11.9026165]
14: 60023 Airbag
[winner: 60023 Airbag, pem_pos=0,
final_scores=-0.9949679, final_local_scores=-0.95491433,
pem=0.0, similarity=-24.406836|
gt: 60023 Airbag, pem_pos=0, final_scores=-0.9949679,
final_local_scores=-0.95491433, pem=0.0,
similarity=-24.406836]
15: 865271 Liver_disease
[winner: 100442 The_Anarchy, pem_pos=0,
final_scores=-1.0973691, final_local_scores=-1.0624549,
pem=0.0, similarity=-27.616062|
RECALL MISS: 865271 Liver_disease]
[winner: 17384301 Liver, pem_pos=0,
final_scores=-1.2279844, final_local_scores=-1.2122371,
pem=-0.05744374, similarity=-31.773472|
gt: 865271 Liver_disease, pem_pos=9,
final_scores=-2.236553, final_local_scores=-2.3220031,
pem=-5.456165, similarity=-35.532017]
16: 6793142 Niigata,_Niigata
[winner: 4560542 City_Hospital_railway_station,
pem_pos=1, final_scores=-0.72778547,
final_local_scores=-0.6232434, pem=-1.2797945,
similarity=-7.5493913|
RECALL MISS: 6793142 Niigata,_Niigata]
[winner: 5324854 Niigata_Transys, pem_pos=4,

final_scores=-1.0679815, final_local_scores=-1.0286534,
pem=-1.9419875, similarity=-16.046526|
gt: 6793142 Niigata, Niigata, pem_pos=0,
final_scores=-1.3575789, final_local_scores=-1.5368333,
pem=-0.5194453, similarity=-38.94763]
17: 34658 1995
[winner: 457278 March, Cambridgeshire, pem_pos=1,
final_scores=-0.629353, final_local_scores=-0.5900179,
pem=-1.4151726, similarity=-5.8216696|
RECALL MISS: 34658 1995]
[winner: 34658 1995, pem_pos=0, final_scores=-0.9944319,
final_local_scores=-0.9616648, pem=-0.68201965,
similarity=-20.899353|
gt: 34658 1995, pem_pos=0, final_scores=-0.9944319,
final_local_scores=-0.9616648, pem=-0.68201965,
similarity=-20.899353]
18: 21710 Nerve_agent
[winner: 5170484 Infraorbital_nerve, pem_pos=0,
final_scores=-0.7796194, final_local_scores=-0.62865454,
pem=0.0, similarity=-14.670592|
gt: 21710 Nerve_agent, pem_pos=3,
final_scores=-1.6285615, final_local_scores=-1.6832776,
pem=-3.9280057, similarity=-24.781542]
[winner: 18993869 Gas, pem_pos=0,
final_scores=-1.2178061, final_local_scores=-1.2204976,
pem=-0.3543779, similarity=-30.405209|
RECALL MISS: 21710 Nerve_agent]
1: 6793142 Niigata, Niigata
[pem_pos=0, final_scores=0.15669997,
final_local_scores=0.14714378, pem=-0.5194453,
similarity=11.305601]
2: 30057 Tokyo
[pem_pos=0, final_scores=0.49486494,
final_local_scores=0.46168745, pem=-0.09283106,
similarity=18.372217]
3: 34267361 Aum_Shinrikyo
[pem_pos=0, final_scores=0.29039747,
final_local_scores=0.36487073, pem=0.0, similarity=14.978184]
1: 15573 Japan
[pem_pos=0, final_scores=0.93017894,
final_local_scores=0.9114807, pem=-0.99076414,
similarity=36.678043]
2: 15573 Japan
[pem_pos=0, final_scores=0.6367144,
final_local_scores=0.5909152, pem=-0.13549899,
similarity=22.460665]
3: 15573 Japan
[pem_pos=0, final_scores=0.79127604,
final_local_scores=0.75975025, pem=-0.13549899,
similarity=27.499039]
4: 30057 Tokyo
[pem_pos=0, final_scores=0.5111967,

```
final_local_scores=0.47952735, pem=-0.09283106,  
similarity=18.904594]  
5: 15573 Japan  
    [pem_pos=0, final_scores=0.68930316,  
final_local_scores=0.64836043, pem=-0.13549899,  
similarity=24.174946]
```

Sample: b'APW19990827_0184_htm'

O'Hare Closes Terminal for Security CHICAGO -LRB- AP -RRB-
-- [1 1 United Airlines 1] shut down a 1 terminal for hours
so authorities could search for a passenger who ran past a 2 s
ecurity checkpoint and disappeared into the crowd at [2 2 O'H
are International Airport 2] . [1 United 1] spokeswoman
Kristina Price said 131 flights were canceled Thursday and
more than 100 were delayed nationwide because of the
incident . About 6,000 passengers were cleared from the
terminal during the search , airline officials said . Eighty
more flights were canceled today and numerous others were
delayed as the effects of the shutdown lingered , Price
said , adding that the carrier hoped to be back to a normal
schedule by this afternoon . The man was not found and [2
Chicago 2] police said officers turned up nothing
suspicious . He was leaving a 3 concourse and heading toward
the 4 baggage claim area when he turned back and ran up a one-
way staircase leading to the concourse , [3 United 3]
spokesman Matthew Triaca said . Guards are stationed at the
staircase , but the man ran past them . No [4 United 4]
flights left the terminal during the search and only a few
landed at other [5 United 5] gates . Some flights resumed by
early evening .

1: 32307 United_Airlines

2: 248449 O'Hare_International_Airport

1: 423849 Airport_terminal

[winner: 8029864 Terminus_Centre-Ville_(AMT), pem_pos=10,
final_scores=-1.261621, final_local_scores=-1.2650254,
pem=-2.4829726, similarity=-20.158377]

gt: 423849 Airport_terminal, pem_pos=1,
final_scores=-1.437207, final_local_scores=-1.5652499,
pem=-1.4373698, similarity=-34.803825]

2: 284770 Airport_security

[winner: 8116691 Security_checkpoint, pem_pos=0,
final_scores=-0.9575993, final_local_scores=-0.93375874,
pem=-0.04800922, similarity=-23.51443]

gt: 284770 Airport_security, pem_pos=2,
final_scores=-1.800053, final_local_scores=-1.919085,
pem=-4.158883, similarity=-30.562958]

3: 1103409 Concourse

[winner: 1440714 IND_Concourse_Line, pem_pos=0,
final_scores=-0.99811983, final_local_scores=-0.990606,
pem=0.0, similarity=-25.471947]

gt: 1103409 Concourse, pem_pos=1,
final_scores=-1.2773526, final_local_scores=-1.3216248,

```

pem=-0.8451948, similarity=-30.753906 ]
4: 973701 Baggage_claim
    [winner: 973701 Baggage_claim, pem_pos=0,
final_scores=-0.9486049, final_local_scores=-0.9798706,
pem=0.0, similarity=-25.15158]
    gt: 973701 Baggage_claim, pem_pos=0,
final_scores=-0.9486049, final_local_scores=-0.9798706,
pem=0.0, similarity=-25.15158 ]
1: 32307 United_Airlines
    [pem_pos=0, final_scores=0.4352548,
final_local_scores=0.38175005, pem=0.0, similarity=15.481896]
2: 248449 O'Hare_International_Airport
    [pem_pos=0, final_scores=0.11146735,
final_local_scores=0.08254714, pem=0.0, similarity=6.5530844]
1: 32307 United_Airlines
    [pem_pos=0, final_scores=0.06242098,
final_local_scores=-0.044903338, pem=-1.5807481,
similarity=11.346062]
2: 6886 Chicago
    [pem_pos=0, final_scores=0.4259675,
final_local_scores=0.4444874, pem=-0.20658566,
similarity=18.477547]
3: 32307 United_Airlines
    [pem_pos=0, final_scores=0.12607642,
final_local_scores=0.036246777, pem=-1.5807481,
similarity=13.767743]
4: 32307 United_Airlines
    [pem_pos=0, final_scores=0.11982258,
final_local_scores=0.029415399, pem=-1.5807481,
similarity=13.563881]
5: 32307 United_Airlines
    [pem_pos=0, final_scores=0.058697537,
final_local_scores=-0.04443142, pem=-1.5807481,
similarity=11.360145]

```

Sample: b'APW19981130_0743_htm'

[1 Portugal 1] and [2 Spain 2] reach agreement on common water supplies [1 1 Portugal 1] and [2 2 Spain 2] on Monday reached an agreement over shared 1 water resources on the [3 3 Iberian 3] peninsula , wrapping up a two-day summit that both premiers called historic . `` It is an ambitious , modern , wide-ranging and efficient agreement , '' [3 Portugal 3] 's 2 Prime Minister [4 4 Antonio Guterres 4] said at the end of the summit in [4 Albufeira 4] , about 200 kilometers -LRB- 125 miles -RRB- south of [5 5 Lisbon 5] . Both governments worked out a shared management plan to monitor water 3 volume in the rivers which supply both countries . [5 Spanish 5] Premier [6 6 Jose Maria Aznar 6] said the agreement reflected `` an exceptional moment '' of relations between the two [6 Iberian 6] countries . 4 Water supply on the southwest [7 European 7] peninsula was one of the few sticking points between the two countries , which have strong trade links and

close political ties . [8 Spanish 8] plans to develop 5 hydro electric projects on its side of the border have been a constant threat to [9 Portugal 9] 's supply , which comes mainly from rivers with sources in the neighboring country . Both governments also reiterated their decision jointly to campaign for continued development funds provided by wealthier [7 7 European Union 7] nations . Cabinet Members also agreed to build new road and railway links between the two countries . The [10 Albufeira 10] summit also produced a plan to increase cooperation between police at the border , in order to curb 6 illegal immigration and 7 drug trafficking .

1: 23033 Portugal
2: 26667 Spain
3: 14883 Iberian_Peninsula
4: 471806 António_Guterres
5: 18091 Lisbon
6: 197618 José_María_Aznar
7: 9317 European_Union

1: 18581463 Water_resources

[winner: 1424697 Ap_(water), pem_pos=12,
final_scores=-1.1881243, final_local_scores=-1.1006236,
pem=-2.7080503, similarity=-14.028291]

RECALL MISS: 18581463 Water_resources]

[winner: 2702905 Resource_(Windows), pem_pos=3,
final_scores=-1.2260319, final_local_scores=-1.1446671,
pem=-1.2756059, similarity=-23.132494]

RECALL MISS: 18581463 Water_resources]

2: 24109 Prime_minister

[winner: 23666 Prime_number, pem_pos=2,
final_scores=-0.59798324, final_local_scores=-0.5062144,
pem=-2.391252, similarity=1.987267]

gt: 24109 Prime_minister, pem_pos=19,
final_scores=-1.0853559, final_local_scores=-1.1480286,
pem=-4.133042, similarity=-7.693629]

[winner: 739237 Minister_(government), pem_pos=0,
final_scores=-1.2133276, final_local_scores=-1.2970223,
pem=-0.49105036, similarity=-31.94561]

RECALL MISS: 24109 Prime_minister]

3: 32498 Volume

[winner: 32498 Volume, pem_pos=0,
final_scores=-1.2837702, final_local_scores=-1.277338,
pem=-0.25805655, similarity=-32.625248]

gt: 32498 Volume, pem_pos=0, final_scores=-1.2837702,
final_local_scores=-1.277338, pem=-0.25805655,
similarity=-32.625248]

4: 684928 Water_supply

[winner: 2651593 Water_(2005_film), pem_pos=1,
final_scores=-0.7108041, final_local_scores=-0.6141974,
pem=-2.2621994, similarity=-1.9369701]

RECALL MISS: 684928 Water_supply]

[winner: 164989 Loss_of_supply, pem_pos=0,

final_scores=-1.468777, final_local_scores=-1.4745686,
pem=-1.2611313, similarity=-33.056126|
gt: 684928 Water_supply, pem_pos=1,
final_scores=-1.7776912, final_local_scores=-1.892901,
pem=-1.2831585, similarity=-45.42021]
5: 381399 Hydroelectricity
[winner: 14073 Hydropower, pem_pos=0,
final_scores=-1.2771466, final_local_scores=-1.3220544,
pem=0.0, similarity=-35.363026|
gt: 381399 Hydroelectricity, pem_pos=1,
final_scores=-1.419718, final_local_scores=-1.4841094,
pem=-0.021431543, similarity=-40.082523]
6: 7124412 Illegal_immigration
[winner: 4016305 Illegal_immigrants_in_Malaysia,
pem_pos=0, final_scores=-0.7575823,
final_local_scores=-0.68978584, pem=0.0,
similarity=-16.494873|
gt: 7124412 Illegal_immigration, pem_pos=1,
final_scores=-0.98114085, final_local_scores=-1.045307,
pem=-0.13563715, similarity=-26.366718]
7: 218449 Illegal_drug_trade
[winner: 803443 ATC_code_G, pem_pos=10,
final_scores=-0.9112594, final_local_scores=-0.6815562,
pem=-3.014498, similarity=0.14402547|
RECALL MISS: 218449 Illegal_drug_trade|
[winner: 19333662 Human_trafficking, pem_pos=0,
final_scores=-1.2260494, final_local_scores=-1.2768803,
pem=0.0, similarity=-34.014942|
gt: 218449 Illegal_drug_trade, pem_pos=1,
final_scores=-1.4686551, final_local_scores=-1.5776691,
pem=-1.3696333, similarity=-35.5428]
1: 23033 Portugal
[pem_pos=0, final_scores=0.60873383,
final_local_scores=0.5585918, pem=-0.30942732,
similarity=22.44192]
2: 26667 Spain
[pem_pos=0, final_scores=0.5881168,
final_local_scores=0.5411839, pem=-0.18857051,
similarity=21.265196]
3: 14883 Iberian_Peninsula
[pem_pos=0, final_scores=0.22146706,
final_local_scores=0.1931538, pem=-0.60556847,
similarity=13.146983]
4: 471806 António_Guterres
[pem_pos=0, final_scores=0.40928656,
final_local_scores=0.4912417, pem=0.0, similarity=18.749346]
5: 18091 Lisbon
[pem_pos=0, final_scores=0.49646753,
final_local_scores=0.47739005, pem=-0.09749677,
similarity=18.866188]
6: 197618 José_María_Aznar
[pem_pos=0, final_scores=0.48464388,

final_local_scores=0.5482294, pem=0.0, similarity=20.449972]
7: 9317 European_Union
[pem_pos=0, final_scores=0.43319935,
final_local_scores=0.4608603, pem=-0.005814796,
similarity=17.874327]
1: 23033 Portugal
[pem_pos=0, final_scores=0.5153613,
final_local_scores=0.45659643, pem=-0.30942732,
similarity=19.398174]
2: 26667 Spain
[pem_pos=0, final_scores=0.52873,
final_local_scores=0.47631276, pem=-0.18857051,
similarity=19.329311]
3: 23033 Portugal
[pem_pos=0, final_scores=0.685332,
final_local_scores=0.6422636, pem=-0.30942732,
similarity=24.938852]
4: 29354765 Albufeira
[pem_pos=0, final_scores=0.24965785,
final_local_scores=0.26484132, pem=0.0, similarity=11.993106]
5: 26667 Spain
[pem_pos=1, final_scores=0.61587995,
final_local_scores=0.5715109, pem=-1.2425895,
similarity=27.902134]
6: 14883 Iberian_Peninsula
[pem_pos=0, final_scores=0.30325693,
final_local_scores=0.28249678, pem=-0.60556847,
similarity=15.813155]
7: 9239 Europe
[pem_pos=0, final_scores=0.54852337,
final_local_scores=0.543884, pem=-0.25557294,
similarity=21.71014]
8: 26667 Spain
[pem_pos=1, final_scores=0.56445223,
final_local_scores=0.5153339, pem=-1.2425895,
similarity=26.225698]
9: 23033 Portugal
[pem_pos=0, final_scores=0.7005629,
final_local_scores=0.65890104, pem=-0.30942732,
similarity=25.435347]
10: 29354765 Albufeira
[pem_pos=0, final_scores=0.21358423,
final_local_scores=0.22543639, pem=0.0, similarity=10.817184]

Sample: b'APW19981022_0710_htm'

[1 1 Swiss 1] government eyes shutting down nuclear power plants The [1 Swiss 1] government Thursday announced that it is planning the country 's eventual exit from 1 nuclear power , saying it wants to support instead 2 hydroelectric plants . But 3 environmental campaigners criticized the lack of a timeframe for closing the country 's 4 four nuclear power plants . Government officials will work with representatives of

the 5 power industry , environmental groups and local governments to discuss a timetable and removal of 6 nuclear waste , a government statement said . It is unrealistic to build more 7 power stations because of environmental concerns , the government has said . Should the discussions between concerned parties fail to produce an agreed timetable , the federal 8 cabinet will make the decision , said the statement . `` We have set out on a new course , '' said Federal Energy Minister [2 2 Moritz Leuenberger 2] . `` Details have yet to be decided . '' In a related move , the federal cabinet ordered energy officials to draw up a discussion paper on changing laws governing the domestic 9 electricity market . The government envisages the creation of a national company to own the power 10 infrastructure . A timeframe for carrying out 11 liberalization of the electricity market itself remains to be determined , the cabinet said . The cabinet said it wants to support hydroelectric plants . It ordered a study of the possibility of 12 tax breaks for hydroelectric plants and how to encourage upgrades of those plants .

1: 26748 Switzerland

2: 155831 Moritz_Leuenberger

1: 22153 Nuclear_power

[winner: 916971 Nuclear_marine_propulsion, pem_pos=1, final_scores=-1.2200745, final_local_scores=-1.2060957, pem=-1.0862324, similarity=-25.995487]

gt: 22153 Nuclear_power, pem_pos=0, final_scores=-1.3422272, final_local_scores=-1.4028647, pem=-0.05378531, similarity=-37.482075]

2: 381399 Hydroelectricity

[winner: 3907148 Generation_III_reactor, pem_pos=4, final_scores=-1.1946981, final_local_scores=-1.1535048, pem=-1.3862944, similarity=-22.794285]

RECALL MISS: 381399 Hydroelectricity]

[winner: 14073 Hydropower, pem_pos=0, final_scores=-1.2312872, final_local_scores=-1.2464507, pem=0.0, similarity=-33.10686]

gt: 381399 Hydroelectricity, pem_pos=1, final_scores=-1.3500459, final_local_scores=-1.3887892, pem=-0.021431543, similarity=-37.23798]

3: 1866009 Environmentalism

[winner: 12230179 Civil_society_campaign, pem_pos=0, final_scores=-1.2998463, final_local_scores=-1.3177247, pem=-0.074108, similarity=-34.83081]

RECALL MISS: 1866009 Environmentalism]

[winner: 3909826 Environmental_planning, pem_pos=6, final_scores=-1.3575251, final_local_scores=-1.3868613, pem=-1.9381839, similarity=-26.756847]

gt: 1866009 Environmentalism, pem_pos=1, final_scores=-1.5179384, final_local_scores=-1.5930142, pem=-1.3395729, similarity=-36.1642]

4: 10878364 Nuclear_power_in_Switzerland

[winner: 7400363 The_Constitution_of_Liberty, pem_pos=1,
final_scores=-1.1386743, final_local_scores=-1.1097469,
pem=-1.6197149, similarity=-20.219086]

RECALL MISS: 10878364 Nuclear_power_in_Switzerland]

[winner: 2087414 TransAlta, pem_pos=0,
final_scores=-1.1882442, final_local_scores=-1.1654727,
pem=0.0, similarity=-30.690315]

RECALL MISS: 10878364 Nuclear_power_in_Switzerland]

[winner: 6235 Cell_nucleus, pem_pos=4,
final_scores=-1.5061136, final_local_scores=-1.4832251,
pem=-2.2086225, similarity=-28.161848]

RECALL MISS: 10878364 Nuclear_power_in_Switzerland]

5: 277750 Electric_power_industry

[winner: 3810208 Power_Company, pem_pos=0,
final_scores=-1.1377624, final_local_scores=-1.0909932,
pem=-0.09097175, similarity=-27.972979]

gt: 277750 Electric_power_industry, pem_pos=1,
final_scores=-1.3296151, final_local_scores=-1.3804493,
pem=-2.730029, similarity=-22.259323]

6: 37257 Radioactive_waste

[winner: 37257 Radioactive_waste, pem_pos=0,
final_scores=-1.1023401, final_local_scores=-1.1224053,
pem=-0.034439277, similarity=-29.217812]

gt: 37257 Radioactive_waste, pem_pos=0,
final_scores=-1.1023401, final_local_scores=-1.1224053,
pem=-0.034439277, similarity=-29.217812]

7: 212141 Power_station

[winner: 3331228 Optical_power, pem_pos=4,
final_scores=-1.3148923, final_local_scores=-1.2363262,
pem=-2.2721138, similarity=-20.448616]

RECALL MISS: 212141 Power_station]

[winner: 15256577 Rowley_(MBTA_station), pem_pos=8,
final_scores=-1.483113, final_local_scores=-1.4303658,
pem=-2.50274, similarity=-24.984968]

RECALL MISS: 212141 Power_station]

8: 32187890 Cabinet_(government)

[winner: 322678 Cabinet_(file_format), pem_pos=12,
final_scores=-1.3014681, final_local_scores=-1.2106698,
pem=-3.543335, similarity=-12.769877]

gt: 32187890 Cabinet_(government), pem_pos=1,
final_scores=-1.7570531, final_local_scores=-1.8644011,
pem=-2.1059742, similarity=-40.09512]

9: 9550 Electricity

[winner: 215909 Electricity_market, pem_pos=0,
final_scores=-0.96770537, final_local_scores=-0.9338595,
pem=-0.0024184003, similarity=-23.765366]

RECALL MISS: 9550 Electricity]

10: 174521 Infrastructure

[winner: 174521 Infrastructure, pem_pos=0,
final_scores=-1.2845256, final_local_scores=-1.3380694,
pem=-0.05070001, similarity=-35.565235]

gt: 174521 Infrastructure, pem_pos=0,

```

final_scores=-1.2845256, final_local_scores=-1.3380694,
pem=-0.05070001, similarity=-35.565235 ]
11: 189004 Liberalization
    [winner: 189004 Liberalization, pem_pos=0,
final_scores=-1.2318335, final_local_scores=-1.2734671,
pem=-0.0995611, similarity=-33.371655]
    gt: 189004 Liberalization, pem_pos=0,
final_scores=-1.2318335, final_local_scores=-1.2734671,
pem=-0.0995611, similarity=-33.371655 ]
12: 2151903 Tax_break
    [winner: 2151903 Tax_break, pem_pos=0,
final_scores=-1.2539529, final_local_scores=-1.2368066,
pem=-0.12131536, similarity=-32.159336]
    gt: 2151903 Tax_break, pem_pos=0,
final_scores=-1.2539529, final_local_scores=-1.2368066,
pem=-0.12131536, similarity=-32.159336 ]
1: 26748 Switzerland
    [pem_pos=0, final_scores=0.77172166,
final_local_scores=0.7795489, pem=-0.1392413,
similarity=28.110222]
2: 155831 Moritz_Leuenberger
    [pem_pos=0, final_scores=0.38340598,
final_local_scores=0.52846426, pem=0.0, similarity=19.86014]
1: 26748 Switzerland
    [pem_pos=0, final_scores=0.80205756,
final_local_scores=0.81268626, pem=-0.1392413,
similarity=29.099108]

```

Sample: b'APW20000303_0067_hm'

[1 The Nation 1] 's Weather 1 Severe thunderstorms were forecast for much of the [1 1 Gulf Coast 1] , from [2 2 Louisiana 2] to [3 3 Georgia 3] and northward to [4 4 Tennessee 4] . 2 Rain was forecast in 5 the [5 Carolinas 5] by tonight . 3 Lightning , large 4 hail , 5 wind of 60 mph and more than 2 inches of rain were possible . Lighter amounts were expected in the [2 6 Ohio Valley 2] . The same storm system roared through [3 Texas 3] and [6 6 Oklahoma 6] on Thursday , hitting areas with wind , hail and heavy rain . Some flights at [7 7 Dallas-Fort Worth International Airport 7] were delayed up to seven hours . In the [8 8 Northeast 8] , as much as 8 inches of sunnyow were possible in the [9 9 Adirondack 9] and [10 10 Green Mountains 10] . sunnyow was forecast from northwesterain [11 11 Pennsylvania 11] into [12 12 Maine 12] . Gusty winds were expected to keep 7 wind-chill readings well below 8 freezing . Mostly quiet conditions were forecast for 9 the West , [13 13 Rockies 13] , 10 Southeast and northerain 11 Plains today . Showers were possible in the Southerain 12 Califorainia and along the coasts of [4 13 Washington 4] and [14 14 Oregon 14] . Highs today were forecast to reach the 30s in [5 New England 5] ; the 40s in the [6 Northeast 6] , [15 15 Midwest 15] , [16 16 Great Lakes 16] and northerain [7 Rockies 7] ; the 50s in 14 Northwest , 15 mi

d-Atlantic states and northern Plains ; and the 60s , 70s and 80s in the Southeast , [8 Southwest 8] and southern Plains . The highest 16 temperature reported in the continental [9 United States 9] on Thursday was 90 degrees in the [10 Texas 10] communities of 17 Cotulla and 18 Alice . The low was minus-4 at [17 17 Grand Marais 17] , [11 Minn 11]

- 1: 11969 Gulf_Coast_of_the_United_States
- 2: 18130 Louisiana
- 3: 48830 Georgia_(U.S._state)
- 4: 30395 Tennessee
- 5: 340152 The_Carolinas
- 6: 22489 Oklahoma
- 7: 248480 Dallas/Fort_Worth_International_Airport
- 8: 431669 Northeastern_United_States
- 9: 54980 Adirondack_Mountains
- 10: 521031 Green_Mountains
- 11: 23332 Pennsylvania
- 12: 19977 Maine
- 13: 25459 Rocky_Mountains
- 14: 26811621 Oregon
- 15: 104697 Midwestern_United_States
- 16: 12010 Great_Lakes
- 17: 119809 Grand_Marais,_Minnesota

1: 70807 Thunderstorm

[winner: 18603166 Raphaël_Sévère, pem_pos=2, final_scores=-0.598187, final_local_scores=-0.47553247, pem=-0.6931472, similarity=-6.3316827]

RECALL MISS: 70807 Thunderstorm]

[winner: 70807 Thunderstorm, pem_pos=0, final_scores=-1.2395492, final_local_scores=-1.3240802, pem=-0.008230489, similarity=-35.378723]

gt: 70807 Thunderstorm, pem_pos=0, final_scores=-1.2395492, final_local_scores=-1.3240802, pem=-0.008230489, similarity=-35.378723]

2: 19009110 Rain

[winner: 1536200 Rain_(entertainer), pem_pos=0, final_scores=-0.57824814, final_local_scores=-0.53828377, pem=-0.8502017, similarity=-7.3502197]

gt: 19009110 Rain, pem_pos=1, final_scores=-0.6253034, final_local_scores=-0.6255736, pem=-1.4920679, similarity=-6.4645534]

3: 61344 Lightning

[winner: 61344 Lightning, pem_pos=0, final_scores=-0.5477899, final_local_scores=-0.53330326, pem=-0.5449503, similarity=-8.861595]

gt: 61344 Lightning, pem_pos=0, final_scores=-0.5477899, final_local_scores=-0.53330326, pem=-0.5449503, similarity=-8.861595]

4: 14458 Hail

[winner: 14458 Hail, pem_pos=1, final_scores=-1.2764853, final_local_scores=-1.3488464, pem=-0.012270081,

similarity=-36.095825|
gt: 14458 Hail, pem_pos=1, final_scores=-1.2764853,
final_local_scores=-1.3488464, pem=-0.012270081,
similarity=-36.095825]
5: 15440316 Wind
[winner: 15440316 Wind, pem_pos=0,
final_scores=-1.251958, final_local_scores=-1.3060119,
pem=-0.43390265, similarity=-32.52466|
gt: 15440316 Wind, pem_pos=0, final_scores=-1.251958,
final_local_scores=-1.3060119, pem=-0.43390265,
similarity=-32.52466]
6: 22388 Ohio_River
[winner: 505873 Ohio_Country, pem_pos=0,
final_scores=0.134795, final_local_scores=0.15265822,
pem=-0.44271705, similarity=11.052902|
gt: 22388 Ohio_River, pem_pos=1,
final_scores=0.120044455, final_local_scores=0.075305074,
pem=-0.9686718, similarity=11.604754]
7: 154910 Wind_chill
[winner: 154910 Wind_chill, pem_pos=0,
final_scores=-1.4189425, final_local_scores=-1.5057077,
pem=0.0, similarity=-40.84361|
gt: 154910 Wind_chill, pem_pos=0,
final_scores=-1.4189425, final_local_scores=-1.5057077,
pem=0.0, similarity=-40.84361]
8: 198824 Freezing
[winner: 198824 Freezing, pem_pos=0,
final_scores=-1.1660999, final_local_scores=-1.147393,
pem=-0.14790457, similarity=-29.346455|
gt: 198824 Freezing, pem_pos=0, final_scores=-1.1660999,
final_local_scores=-1.147393, pem=-0.14790457,
similarity=-29.346455]
9: 179578 Western_United_States
[winner: 179578 Western_United_States, pem_pos=2,
final_scores=-0.18578787, final_local_scores=-0.2798579,
pem=-2.7524204, similarity=10.706285|
gt: 179578 Western_United_States, pem_pos=2,
final_scores=-0.18578787, final_local_scores=-0.2798579,
pem=-2.7524204, similarity=10.706285]
[winner: 3349493 Male-
female_income_disparity_in_the_United_States, pem_pos=0,
final_scores=-1.5307041, final_local_scores=-1.5734005,
pem=-0.12731102, similarity=-42.171356|
RECALL MISS: 179578 Western_United_States]
10: 816925 Southeastern_United_States
[winner: 28741 Southeast_Asia, pem_pos=4,
final_scores=-0.12530358, final_local_scores=-0.12543866,
pem=-2.268348, similarity=12.682002|
gt: 816925 Southeastern_United_States, pem_pos=0,
final_scores=-0.13480376, final_local_scores=-0.19882596,
pem=-1.5138595, similarity=6.388953]
11: 51464 Great_Plains

[winner: 51464 Great_Plains, pem_pos=3,
final_scores=-0.08782606, final_local_scores=-0.14059812,
pem=-1.7705977, similarity=9.522771|
gt: 51464 Great_Plains, pem_pos=3,
final_scores=-0.08782606, final_local_scores=-0.14059812,
pem=-1.7705977, similarity=9.522771]
12: 5407 California no overlap with any filtered span
13: 13015878 Washington_(state)
[winner: 108956 Washington,_D.C., pem_pos=0,
final_scores=0.2399617, final_local_scores=0.21036041,
pem=-1.7496481, similarity=19.882145|
gt: 13015878 Washington_(state), pem_pos=1,
final_scores=0.025702015, final_local_scores=-0.047689915,
pem=-1.9417106, similarity=13.225875]
14: 78147 Pacific_Northwest
[winner: 146068 Northwest_Airlines, pem_pos=3,
final_scores=-0.22118612, final_local_scores=-0.2094625,
pem=-2.3414865, similarity=10.572302|
gt: 78147 Pacific_Northwest, pem_pos=8,
final_scores=-0.30785608, final_local_scores=-0.39927104,
pem=-2.8211143, similarity=7.5163264]
15: 181779 Mid-Atlantic_states
[winner: 181779 Mid-Atlantic_states, pem_pos=0,
final_scores=-0.78036517, final_local_scores=-0.8976362,
pem=0.0, similarity=-22.69754|
gt: 181779 Mid-Atlantic_states, pem_pos=0,
final_scores=-0.78036517, final_local_scores=-0.8976362,
pem=0.0, similarity=-22.69754]
[winner: 285512 State_(computer_science), pem_pos=1,
final_scores=-1.4036984, final_local_scores=-1.3744934,
pem=-2.1914163, similarity=-25.01065|
RECALL MISS: 181779 Mid-Atlantic_states]
16: 20647050 Temperature
[winner: 316737 Autoignition_temperature, pem_pos=3,
final_scores=-1.2865092, final_local_scores=-1.2457333,
pem=-1.0986123, similarity=-27.111025|
gt: 20647050 Temperature, pem_pos=0,
final_scores=-1.2881453, final_local_scores=-1.3259978,
pem=-0.03605954, similarity=-35.284607]
17: 136261 Cotulla,_Texas
[winner: 136261 Cotulla,_Texas, pem_pos=0,
final_scores=0.028045878, final_local_scores=0.048635572,
pem=0.0, similarity=5.5410953|
gt: 136261 Cotulla,_Texas, pem_pos=0,
final_scores=0.028045878, final_local_scores=0.048635572,
pem=0.0, similarity=5.5410953]
18: 136209 Alice,_Texas
[winner: 136209 Alice,_Texas, pem_pos=3,
final_scores=-0.4055642, final_local_scores=-0.43961185,
pem=-2.7932734, similarity=6.161073|
gt: 136209 Alice,_Texas, pem_pos=3,
final_scores=-0.4055642, final_local_scores=-0.43961185,

pem=-2.7932734, similarity=6.161073]

1: 11969 Gulf_Coast_of_the_United_States

[pem_pos=0, final_scores=0.15001164,
final_local_scores=0.11652464, pem=-0.19244206,
similarity=8.61357]

2: 18130 Louisiana

[pem_pos=0, final_scores=0.37003547,
final_local_scores=0.3316558, pem=-0.17745006,
similarity=14.951984]

3: 48830 Georgia_(U.S._state)

[pem_pos=0, final_scores=0.3150429,
final_local_scores=0.26043615, pem=-0.70047224,
similarity=15.670923]

4: 30395 Tennessee

[pem_pos=0, final_scores=0.45314664,
final_local_scores=0.41407365, pem=-0.27977526,
similarity=17.967958]

5: 340152 The_Carolinas

[pem_pos=0, final_scores=0.20152967,
final_local_scores=0.1662589, pem=-0.04036205,
similarity=9.270705]

6: 22489 Oklahoma

[pem_pos=0, final_scores=0.5957076,
final_local_scores=0.56377345, pem=-0.35266355,
similarity=22.831676]

7: 248480 Dallas/Fort_Worth_International_Airport

[pem_pos=0, final_scores=0.20156495,
final_local_scores=0.23007223, pem=0.0, similarity=10.955527]

8: 431669 Northeastern_United_States

[pem_pos=0, final_scores=0.08028914,
final_local_scores=0.025345683, pem=-0.83109236,
similarity=9.365688]

9: 54980 Adirondack_Mountains

[pem_pos=1, final_scores=0.110276446,
final_local_scores=0.07381953, pem=-0.95946366,
similarity=11.510347]

10: 521031 Green_Mountains

[pem_pos=0, final_scores=0.35271567,
final_local_scores=0.3843431, pem=0.0, similarity=15.559278]

11: 23332 Pennsylvania

[pem_pos=0, final_scores=0.52759093,
final_local_scores=0.49344963, pem=-0.11810702,
similarity=19.457518]

12: 19977 Maine

[pem_pos=0, final_scores=0.13395049,
final_local_scores=0.06596008, pem=-0.2749413,
similarity=7.553267]

13: 25459 Rocky_Mountains

[pem_pos=0, final_scores=0.112812325,
final_local_scores=0.066964924, pem=-0.36690158,
similarity=8.083347]

14: 26811621 Oregon

[pem_pos=0, final_scores=0.5705449,
final_local_scores=0.5486938, pem=-0.28686622,
similarity=22.023851]
15: 104697 Midwestern_United_States
[pem_pos=0, final_scores=0.36871272,
final_local_scores=0.32216537, pem=-0.031208755,
similarity=13.873488]
16: 12010 Great_Lakes
[pem_pos=0, final_scores=0.307738,
final_local_scores=0.28820723, pem=-0.09949329,
similarity=13.231452]
17: 119809 Grand_Marais_Minnesota
[pem_pos=1, final_scores=0.27740043,
final_local_scores=0.32790267, pem=-0.037508648,
similarity=14.07896]
1: 39902584 The_Nation
[pem_pos=0, final_scores=0.22798507,
final_local_scores=0.30257204, pem=-0.015505094,
similarity=13.203385]
2: 505873 Ohio_Country
[pem_pos=0, final_scores=0.134795,
final_local_scores=0.15265822, pem=-0.44271705,
similarity=11.052902]
3: 29810 Texas
[pem_pos=0, final_scores=0.698258,
final_local_scores=0.6848047, pem=-0.16191302,
similarity=25.406158]
4: 108956 Washington_D.C.
[pem_pos=0, final_scores=0.2399617,
final_local_scores=0.21036041, pem=-1.7496481,
similarity=19.882145]
5: 21531764 New_England
[pem_pos=0, final_scores=0.3784837,
final_local_scores=0.33618522, pem=-0.13270761,
similarity=14.843835]
6: 431669 Northeastern_United_States
[pem_pos=0, final_scores=0.13652937,
final_local_scores=0.086779565, pem=-0.83109236,
similarity=11.1989975]
7: 25459 Rocky_Mountains
[pem_pos=0, final_scores=0.14898787,
final_local_scores=0.106481194, pem=-0.36690158,
similarity=9.262591]
8: 179571 Southwestern_United_States
[pem_pos=0, final_scores=0.047258094,
final_local_scores=0.0022796988, pem=-0.95406055,
similarity=9.346073]
9: 3434750 United_States
[pem_pos=0, final_scores=0.46561015,
final_local_scores=0.44736075, pem=-0.073457584,
similarity=17.839325]
10: 29810 Texas


```
[pem_pos=0, final_scores=0.64805824,
final_local_scores=0.6299691, pem=-0.16191302,
similarity=23.769756]
11: 19590 Minnesota
[pem_pos=0, final_scores=0.58065104,
final_local_scores=0.5595238, pem=0.0, similarity=20.78702]
```

Sample: b'APW19980903_1073_htm'

[1 Austrian 1] president tells [2 Albright 2] of effort to identify 1 stolen art [1 1 Austria 1] 's president told 2 Secretary of State [2 2 Madeleine Albright 2] on Thursday that the government was establishing a commission to identify art which was stolen from [3 3 Holocaust 3] victims by the [4 4 Nazis 4] . [3 Austrian 3] officials said President [5 5 Thomas Klestil 5] told [4 Albright 4] , who paid a one-day visit to [5 Vienna 5] after the [6 6 Moscow 6] summit , that the commission would first determine the origin of all art works owned by the Federal Museum here . There was no statement on the art issue by [6 U.S. 6] officials . [7 Albright 7] left for the [7 7 United States 7] after meetings with [8 Klestil 8] , Chancellor [8 8 Viktor Klima 8] and Foreign Minister [9 9 Wolfgang Schuessel 9] . [9 Austria 9] , the [10 United States 10] and 37 other nations have launched a drive to identify billions of dollars worth of art stolen from [11 Holocaust 11] victims throughout [12 Europe 12] . The [10 10 World Jewish Congress 10] estimates the current value of stolen art at between dlrs 90 billion and dlrs 140 billion . In June , the 3 Association of Art Museum Directors in the [13 United States 13] recommended that [14 American 14] art museums check their own collections to determine if any works were stolen during the war and set up a worldwide database for tracking . [15 Austria 15] was absorbed into [11 11 Adolf Hitler 11] 's [12 12 Germany 12] in 1938 and was restored as an independent nation following [16 World War II 16] . rr

```
1: 26964606 Austria
2: 133249 Madeleine_Albright
3: 10396793 The_Holocaust
4: 31045316 Nazism
5: 242317 Thomas_Klestil
6: 19004 Moscow
7: 3434750 United_States
8: 299392 Viktor_Klima
9: 289453 Wolfgang_Schuessel
10: 311160 World_Jewish_Congress
11: 2731583 Adolf_Hitler
12: 11867 Germany
```

```
1: 425514 Art_theft
```

```
[winner: 26187979 Stolen_(2009_drama_film), pem_pos=3,
final_scores=-1.1625258, final_local_scores=-1.0841453,
pem=-1.3341784, similarity=-21.007874]
```

```
gt: 425514 Art_theft, pem_pos=5,
final_scores=-1.6345186, final_local_scores=-1.7486277,
```

pem=-1.9459101, similarity=-37.510662]
[winner: 752 Art, pem_pos=0, final_scores=-1.5388346,
final_local_scores=-1.6582763, pem=-0.22415501,
similarity=-44.17757]
RECALL MISS: 425514 Art_theft]
2: 26422038 Secretary_of_State
[winner: 32293 United_States_Secretary_of_State,
pem_pos=0, final_scores=-0.6683297,
final_local_scores=-0.75642425, pem=-0.88363016,
similarity=-13.678178]
gt: 26422038 Secretary_of_State, pem_pos=1,
final_scores=-0.7612054, final_local_scores=-0.8649782,
pem=-1.175439, similarity=-15.330741]
3: 13460890 Association_of_Art_Museum_Directors
[winner: 13460890 Association_of_Art_Museum_Directors,
pem_pos=0, final_scores=0.014554456,
final_local_scores=0.11403814, pem=0.0, similarity=7.492839]
gt: 13460890 Association_of_Art_Museum_Directors,
pem_pos=0, final_scores=0.014554456,
final_local_scores=0.11403814, pem=0.0, similarity=7.492839]
1: 26964606 Austria
[pem_pos=0, final_scores=0.5964216,
final_local_scores=0.5815942, pem=-0.25262418,
similarity=22.819454]
2: 133249 Madeleine_Albright
[pem_pos=0, final_scores=0.59989846,
final_local_scores=0.6266551, pem=0.0, similarity=22.790352]
3: 10396793 The_Holocaust
[pem_pos=0, final_scores=0.30114937,
final_local_scores=0.2830738, pem=0.0, similarity=12.5372]
4: 31045316 Nazism
[pem_pos=0, final_scores=0.07691704,
final_local_scores=0.04223007, pem=-0.009788822,
similarity=5.4031754]
5: 242317 Thomas_Klestil
[pem_pos=0, final_scores=0.9296319,
final_local_scores=1.0182917, pem=0.0, similarity=34.477566]
6: 19004 Moscow
[pem_pos=0, final_scores=0.43043822,
final_local_scores=0.42959976, pem=-0.106582455,
similarity=17.489439]
7: 3434750 United_States
[pem_pos=0, final_scores=0.48960805,
final_local_scores=0.4774117, pem=-0.073457584,
similarity=18.736105]
8: 299392 Viktor_Klima
[pem_pos=0, final_scores=0.726667,
final_local_scores=0.83567655, pem=0.0, similarity=29.027967]
9: 289453 Wolfgang_Schüssel
[pem_pos=0, final_scores=0.7885893,
final_local_scores=0.8794293, pem=0.0, similarity=30.333637]
10: 311160 World_Jewish_Congress

[pem_pos=0, final_scores=0.12873761,
final_local_scores=0.120661914, pem=0.0, similarity=7.690505]
11: 2731583 Adolf_Hitler
[pem_pos=0, final_scores=0.4124034,
final_local_scores=0.40588546, pem=0.0, similarity=16.202145]
12: 11867 Germany
[pem_pos=0, final_scores=0.6028787,
final_local_scores=0.5963748, pem=-0.15529114,
similarity=22.731224]
1: 26964606 Austria
[pem_pos=0, final_scores=0.7088072,
final_local_scores=0.70435834, pem=-0.40674555,
similarity=27.321115]
2: 133249 Madeleine_Albright
[pem_pos=0, final_scores=0.340015,
final_local_scores=0.34954762, pem=0.0, similarity=14.520911]
3: 26964606 Austria
[pem_pos=0, final_scores=0.66254884,
final_local_scores=0.65382814, pem=-0.40674555,
similarity=25.813194]
4: 133249 Madeleine_Albright
[pem_pos=0, final_scores=0.40790766,
final_local_scores=0.42371005, pem=0.0, similarity=16.734066]
5: 55866 Vienna
[pem_pos=0, final_scores=0.6159867,
final_local_scores=0.60465944, pem=-0.12517071,
similarity=22.814653]
6: 3434750 United_States
[pem_pos=0, final_scores=0.53736955,
final_local_scores=0.5295839, pem=-0.14075686,
similarity=20.659008]
7: 133249 Madeleine_Albright
[pem_pos=0, final_scores=0.65811026,
final_local_scores=0.6970184, pem=0.0, similarity=24.890131]
8: 242317 Thomas_Klestil
[pem_pos=0, final_scores=0.36543435,
final_local_scores=0.41413063, pem=0.0, similarity=16.448196]
9: 26964606 Austria
[pem_pos=0, final_scores=0.5336972,
final_local_scores=0.5130773, pem=-0.25262418,
similarity=20.774773]
10: 3434750 United_States
[pem_pos=0, final_scores=0.5339572,
final_local_scores=0.52585644, pem=-0.073457584,
similarity=20.181791]
11: 10396793 The_Holocaust
[pem_pos=0, final_scores=0.34849793,
final_local_scores=0.33479488, pem=0.0, similarity=14.08066]
12: 9239 Europe
[pem_pos=0, final_scores=0.61710757,
final_local_scores=0.6577056, pem=-0.086338006,
similarity=24.18648]

13: 3434750 United_States
 [pem_pos=0, final_scores=0.5133397,
 final_local_scores=0.5033349, pem=-0.073457584,
 similarity=19.509705]
 14: 3434750 United_States
 [pem_pos=0, final_scores=0.59392214,
 final_local_scores=0.59135896, pem=-0.12994486,
 similarity=22.443705]
 15: 26964606 Austria
 [pem_pos=0, final_scores=0.5040351,
 final_local_scores=0.480676, pem=-0.25262418,
 similarity=19.807854]
 16: 32927 World_War_II
 [pem_pos=0, final_scores=0.4321404,
 final_local_scores=0.4200244, pem=0.0, similarity=16.624079]

Sample: b'APW19981001_0866_htm'

[1 Hungarian 1] government and [2 Jewish 2] communities agree on compensation The [1 1 Hungarian 1] government and the Federation of 1 Hungarian Jewish Communities signed an agreement Thursday on collective compensation for 2 communist - era confiscations , the state news agency [2 2 MTI 2] reported . The financial basis for the funding is real estate that belonged to the [3 Jewish 3] communities and fell victim to 3 nationalization under the communists . Instead of getting back the actual property _ 152 schools and other buildings _ the assessed value of which is 13,511,000,000 4 fo rints -LRB- dlrs 63 million -RRB- , the [4 Jewish 4] community collectively is to get a percentage annually , starting with the current year , when 608 million forints - LRB- dlrs 2.9 million -RRB- will be paid . `` The money will be mainly used for education , religious services , 5 social w elfare , health service , and charity ends , '' Gusztav Zoltai , executive director of the Federation , said . There are between 80,000 and 100,000 [5 Jews 5] in [6 Hungary 6] . Some 600,000 were killed in 3 the [3 Holocaust 3] . When the communists assumed full power in 1948 , they started a wave of nationalizations that spared no one , including the [7 Jewish 7] community . On June 20 , 1997 , a similar agreement , the first of its kind , was signed between the [8 Hungarian 8] government and the [4 4 Vatican 4] . -LRB- ab/rp -RRB-

1: 13275 Hungary
 2: 8946055 Magyar_Távirati_Iroda
 3: 10396793 The_Holocaust
 4: 13393 Holy_See

1: 2378869 History_of_the_Jews_in_Hungary
 [winner: 2976271 Italian_Jews, pem_pos=0,
 final_scores=-0.5172979, final_local_scores=-0.5695487,
 pem=0.0, similarity=-12.906756]
 RECALL MISS: 2378869 History_of_the_Jews_in_Hungary]
 [winner: 2749227 Hungarian_folk_music, pem_pos=10,

final_scores=-1.6476699, final_local_scores=-1.695857,
 pem=-3.733428, similarity=-26.215078]
 RECALL MISS: 2378869 History_of_the_Jews_in_Hungary]
 2: 9209651 Communism
 [winner: 9209651 Communism, pem_pos=0,
 final_scores=-1.3554032, final_local_scores=-1.51689,
 pem=-0.08343673, similarity=-40.723568]
 gt: 9209651 Communism, pem_pos=0,
 final_scores=-1.3554032, final_local_scores=-1.51689,
 pem=-0.08343673, similarity=-40.723568]
 3: 1062429 Nationalization
 [winner: 1062429 Nationalization, pem_pos=0,
 final_scores=-1.277832, final_local_scores=-1.3571048,
 pem=-0.019124912, similarity=-36.304996]
 gt: 1062429 Nationalization, pem_pos=0,
 final_scores=-1.277832, final_local_scores=-1.3571048,
 pem=-0.019124912, similarity=-36.304996]
 4: 229795 Hungarian_forint
 [winner: 3359939 Austro-Hungarian_gulden, pem_pos=1,
 final_scores=-1.4405155, final_local_scores=-1.5304472,
 pem=-0.66860604, similarity=-37.9459]
 gt: 229795 Hungarian_forint, pem_pos=0,
 final_scores=-1.4957359, final_local_scores=-1.5911536,
 pem=0.0, similarity=-43.393482]
 5: 247817 Welfare
 [winner: 247817 Welfare, pem_pos=0,
 final_scores=-1.2265404, final_local_scores=-1.3154166,
 pem=-0.16035746, similarity=-34.29289]
 gt: 247817 Welfare, pem_pos=0, final_scores=-1.2265404,
 final_local_scores=-1.3154166, pem=-0.16035746,
 similarity=-34.29289]
 1: 13275 Hungary
 [pem_pos=1, final_scores=0.59615725,
 final_local_scores=0.58815634, pem=-1.0319586,
 similarity=27.253424]
 2: 8946055 Magyar_Távirati_Iroda
 [pem_pos=0, final_scores=0.2184581,
 final_local_scores=0.31794614, pem=-0.5798185,
 similarity=16.731003]
 3: 10396793 The_Holocaust
 [pem_pos=0, final_scores=0.19967337,
 final_local_scores=0.15878823, pem=0.0, similarity=8.828271]
 4: 13393 Holy_See
 [pem_pos=0, final_scores=0.05225362,
 final_local_scores=0.053516626, pem=-0.6812086,
 similarity=9.391274]
 1: 13275 Hungary
 [pem_pos=1, final_scores=0.69415766,
 final_local_scores=0.6952068, pem=-1.0319586,
 similarity=30.448023]
 2: 25955086 Jews
 [pem_pos=1, final_scores=0.3816049,

final_local_scores=0.32344317, pem=-0.80476934,
 similarity=18.118362]
 3: 25955086 Jews
 [pem_pos=1, final_scores=0.50270045,
 final_local_scores=0.45572168, pem=-0.80476934,
 similarity=22.065817]
 4: 25955086 Jews
 [pem_pos=1, final_scores=0.40864462,
 final_local_scores=0.35297996, pem=-0.80476934,
 similarity=18.9998]
 5: 25955086 Jews
 [pem_pos=0, final_scores=0.49547493,
 final_local_scores=0.4478289, pem=-0.1252217,
 similarity=18.134794]
 6: 13275 Hungary
 [pem_pos=0, final_scores=0.6090451,
 final_local_scores=0.6022343, pem=-0.32577527,
 similarity=23.833202]
 7: 25955086 Jews
 [pem_pos=1, final_scores=0.48263496,
 final_local_scores=0.43380314, pem=-0.80476934,
 similarity=21.411724]
 8: 13275 Hungary
 [pem_pos=1, final_scores=0.64836985,
 final_local_scores=0.6451906, pem=-1.0319586,
 similarity=28.955439]

Sample: b'APW19981113_0729_htm'

Airport ready for operation ; [1 Palestinians 1] await word
 from [2 Israel 2] The new [3 1 Palestinian 3] 2 airport , a
 powerful symbol of fledgling 3 independence , is ready to
 open , the head of the Palestinian 4 Civil Aviation Authority
 said Friday after an operations drill with [4 5 Israeli 4]
 officials . [1 1 Yasser Arafat International Airport 1] in
 the southern [2 2 Gaza Strip 2] was to have opened last week
 as part of the Israeli-Palestinian peace agreement .
 However , [5 Israel 5] fell behind schedule when it held up 6
 Cabinet 7 ratification of the accord by two weeks . Prime
 Minister [3 3 Benjamin Netanyahu 3] has said the airport
 would start operating next week , but has not set a date . On
 Thursday , [6 Israeli 6] and [7 Palestinian 7] officials
 staged a four-hour drill , said Fayez Zeidan , head of the
 Palestinian 8 Civil Aviation Authority . About 100 [8
 Palestinian 8] airport employees and 50 [9 9 Israelis 9] who
 will share security duties played the roles of passengers and
 airport personnel . For lack of a plane , the `` passengers
 ' ' arrived by bus . Zeidan said the exercise covered arrival
 and departure procedures , including 10 passport checks as
 well as 11 customs and luggage inspection . He said the
 airport was ready for operation . `` We just need to hear
 from the [10 Israelis 10] , ' ' Zeidan said .
 1: 40390174 Yasser_Arafat_International_Airport

2: 12047 Gaza_Strip
3: 56472 Benjamin_Netanyahu
1: 23267 Palestinian_people
[winner: 241405 State_of_Palestine, pem_pos=9,
final_scores=0.330158, final_local_scores=0.279707,
pem=-2.8508017, similarity=27.93983|
gt: 23267 Palestinian_people, pem_pos=0,
final_scores=-0.34803212, final_local_scores=-0.45787773,
pem=-0.38816506, similarity=-7.4633694]
2: 37575 Airport
[winner: 37575 Airport, pem_pos=0,
final_scores=-1.349098, final_local_scores=-1.3487384,
pem=-0.121571906, similarity=-35.498207|
gt: 37575 Airport, pem_pos=0, final_scores=-1.349098,
final_local_scores=-1.3487384, pem=-0.121571906,
similarity=-35.498207]
3: 62450 Independence
[winner: 22541169
Independence,_Calaveras_County,_California, pem_pos=19,
final_scores=-1.4137286, final_local_scores=-1.3501184,
pem=-3.8581288, similarity=-15.21941|
gt: 62450 Independence, pem_pos=0,
final_scores=-1.4156911, final_local_scores=-1.5470457,
pem=-1.3882917, similarity=-34.52747]
4: 18575446 Civil_aviation
[winner: 20778386 National_aviation_authority, pem_pos=0,
final_scores=-0.56266904, final_local_scores=-0.53578997,
pem=-0.40803733, similarity=-9.680359|
RECALL MISS: 18575446 Civil_aviation]
5: 181296 Israelis
[winner: 9282173 Israel, pem_pos=0,
final_scores=0.5646138, final_local_scores=0.5109474,
pem=-0.34790626, similarity=21.229372|
gt: 181296 Israelis, pem_pos=1, final_scores=-0.3500607,
final_local_scores=-0.40807176, pem=-1.7488571,
similarity=1.4225947]
6: 5006115 Cabinet_of_Israel
[winner: 32187890 Cabinet_(government), pem_pos=1,
final_scores=-0.36373323, final_local_scores=-0.3617466,
pem=-2.1059742, similarity=4.747087|
gt: 5006115 Cabinet_of_Israel, pem_pos=7,
final_scores=-0.48372138, final_local_scores=-0.5785614,
pem=-3.3819857, similarity=5.2160487]
7: 429650 Ratification
[winner: 429650 Ratification, pem_pos=0,
final_scores=-1.3926846, final_local_scores=-1.4709609,
pem=-0.1684445, similarity=-38.890667|
gt: 429650 Ratification, pem_pos=0,
final_scores=-1.3926846, final_local_scores=-1.4709609,
pem=-0.1684445, similarity=-38.890667]
8: 20778386 National_aviation_authority
[winner: 684737

Civil_Aviation_Authority_(United_Kingdom), pem_pos=1,
final_scores=-0.60353494, final_local_scores=-0.5755401,
pem=-0.8439734, similarity=-8.495894|
gt: 20778386 National_aviation_authority, pem_pos=0,
final_scores=-0.6502247, final_local_scores=-0.6314312,
pem=-0.40803733, similarity=-12.534483]
9: 181296 Israelis
[winner: 9282173 Israel, pem_pos=1,
final_scores=0.3197682, final_local_scores=0.24349076,
pem=-0.5460739, similarity=14.325598|
gt: 181296 Israelis, pem_pos=0,
final_scores=-0.0048480034, final_local_scores=-0.030979343,
pem=-0.009432555, similarity=3.2165232]
10: 23287 Passport
[winner: 23287 Passport, pem_pos=0,
final_scores=-1.2565435, final_local_scores=-1.3610239,
pem=-0.067494445, similarity=-36.15891|
gt: 23287 Passport, pem_pos=0, final_scores=-1.2565435,
final_local_scores=-1.3610239, pem=-0.067494445,
similarity=-36.15891]
11: 59214 Customs
[winner: 11377077 Zollfahndungsamt, pem_pos=7,
final_scores=-0.93916476, final_local_scores=-0.8632723,
pem=-2.483149, similarity=-8.168304|
gt: 59214 Customs, pem_pos=0, final_scores=-1.3250124,
final_local_scores=-1.4005833, pem=-0.49704432,
similarity=-35.003483]
1: 40390174 Yasser_Arafat_International_Airport
[pem_pos=0, final_scores=0.16710518,
final_local_scores=0.22802886, pem=0.0, similarity=10.894549]
2: 12047 Gaza_Strip
[pem_pos=0, final_scores=0.39634073,
final_local_scores=0.36645132, pem=0.0, similarity=15.025352]
3: 56472 Benjamin_Netanyahu
[pem_pos=0, final_scores=0.6651851,
final_local_scores=0.6865397, pem=0.0, similarity=24.577427]
1: 241405 State_of_Palestine
[pem_pos=5, final_scores=0.27951074,
final_local_scores=0.22438258, pem=-1.7909594,
similarity=20.525253]
2: 9282173 Israel
[pem_pos=0, final_scores=0.5141452,
final_local_scores=0.45581812, pem=-0.21279564,
similarity=18.84945]
3: 241405 State_of_Palestine
[pem_pos=9, final_scores=0.330158,
final_local_scores=0.279707, pem=-2.8508017,
similarity=27.93983]
4: 9282173 Israel
[pem_pos=0, final_scores=0.5646138,
final_local_scores=0.5109474, pem=-0.34790626,
similarity=21.229372]

5: 9282173 Israel
 [pem_pos=0, final_scores=0.6459466,
 final_local_scores=0.5997911, pem=-0.21279564,
 similarity=23.145893]
 6: 9282173 Israel
 [pem_pos=0, final_scores=0.54169804,
 final_local_scores=0.48591536, pem=-0.34790626,
 similarity=20.482367]
 7: 241405 State_of_Palestine
 [pem_pos=9, final_scores=0.67617553,
 final_local_scores=0.6576786, pem=-2.8508017,
 similarity=39.219257]
 8: 241405 State_of_Palestine
 [pem_pos=9, final_scores=0.30452293,
 final_local_scores=0.25170463, pem=-2.8508017,
 similarity=27.104185]
 9: 9282173 Israel
 [pem_pos=1, final_scores=0.3197682,
 final_local_scores=0.24349076, pem=-0.5460739,
 similarity=14.325598]
 10: 9282173 Israel
 [pem_pos=1, final_scores=0.29522306,
 final_local_scores=0.21667889, pem=-0.5460739,
 similarity=13.525478]

Sample: b'APW19980709_0263_htm'

[1 Vietnam 1] 's prime minister demands improvements by state 1 monopoly 2 Prime Minister [1 1 Phan Van Khai 1] has taken the state 3 electricity monopoly to task for abuse of power and excessive 4 bureaucracy while accusing it of overcharging , the newspaper Dau Tu -LRB- Investment -RRB- reported Thursday . In a blunt assessment , he ordered Electricity of 5 Vietnam to help cut waste that is driving up costs and could spark public discontent . EVN must work out corrective measures based on his instructions , and individual leaders have to do 6 self-criticisms . [2 Khai 2] said the utility has plenty of problems , particularly in the tender process and 7 project accounting , that have led to complaints and affect [3 Vietnam 3] 's reputation with foreign investors . In March , former President [4 Le Duc Anh 4] , in an article for the Army newspaper , criticized irregularities in the bidding process for the country 's largest thermal electric plant . Local media have reported that the company has overblown the costs for construction and repair work on a number of projects to justify overcharging for electricity . [5 Vietnam 5] is chronically short of power , and several projects to raise its output are bogged down in bureaucracy . EVN imposed 8 rotating power cuts earlier this year as the worst 9 drought in a century dropped water levels to critical levels at the 10 reservoir s that feed key 11 hydroelectric plants .

1: 322581 Phan_Văn_Khai

1: 18878 Monopoly
[winner: 2456661 Monopolylogue, pem_pos=9,
final_scores=-1.0578653, final_local_scores=-0.8951587,
pem=-2.5649493, similarity=-8.675016|
gt: 18878 Monopoly, pem_pos=1, final_scores=-1.2663038,
final_local_scores=-1.280865, pem=-1.6982388,
similarity=-24.89857]
2: 24109 Prime_minister
[winner: 23666 Prime_number, pem_pos=2,
final_scores=-1.0174038, final_local_scores=-0.9339554,
pem=-2.391252, similarity=-10.777378|
gt: 24109 Prime_minister, pem_pos=19,
final_scores=-1.286816, final_local_scores=-1.3399613,
pem=-4.133042, similarity=-13.42128]
[winner: 30876704 Minister_(Catholic_Church), pem_pos=9,
final_scores=-1.5883068, final_local_scores=-1.5527353,
pem=-4.768517, similarity=-16.31507|
RECALL MISS: 24109 Prime_minister]
3: 9550 Electricity
[winner: 2144978 Nashville_Electric_Service, pem_pos=2,
final_scores=-1.0265687, final_local_scores=-0.9694059,
pem=-1.0984836, similarity=-18.865568|
gt: 9550 Electricity, pem_pos=0,
final_scores=-1.0683941, final_local_scores=-1.057044,
pem=-0.10830185, similarity=-26.865625]
4: 20949522 Bureaucracy
[winner: 20949522 Bureaucracy, pem_pos=0,
final_scores=-1.3114234, final_local_scores=-1.3368444,
pem=-0.052614767, similarity=-35.51826|
gt: 20949522 Bureaucracy, pem_pos=0,
final_scores=-1.3114234, final_local_scores=-1.3368444,
pem=-0.052614767, similarity=-35.51826]
5: 202354 Vietnam
[winner: 19998253 Vietnam_Electricity, pem_pos=0,
final_scores=-0.018127277, final_local_scores=0.06668164,
pem=0.0, similarity=6.079626|
RECALL MISS: 202354 Vietnam]
6: 2396733 Self-criticism
[winner: 2396733 Self-criticism, pem_pos=0,
final_scores=-0.9658517, final_local_scores=-0.96849906,
pem=0.0, similarity=-24.812231|
gt: 2396733 Self-criticism, pem_pos=0,
final_scores=-0.9658517, final_local_scores=-0.96849906,
pem=0.0, similarity=-24.812231]
7: 912142 Project_accounting
[winner: 912142 Project_accounting, pem_pos=0,
final_scores=-1.0708224, final_local_scores=-0.9870148,
pem=0.0, similarity=-25.364777|
gt: 912142 Project_accounting, pem_pos=0,
final_scores=-1.0708224, final_local_scores=-0.9870148,
pem=0.0, similarity=-25.364777]
8: 296636 Power_outage

[winner: 15916022 Cuts,_Oise, pem_pos=1,
 final_scores=-1.0702504, final_local_scores=-0.9692512,
 pem=-1.0432537, similarity=-19.161303|
 RECALL MISS: 296636 Power_outage]
 [winner: 39789 Rotation, pem_pos=0,
 final_scores=-1.1683004, final_local_scores=-1.094605,
 pem=-0.040590763, similarity=-28.354744|
 RECALL MISS: 296636 Power_outage]
 [winner: 3331228 Optical_power, pem_pos=4,
 final_scores=-1.2773062, final_local_scores=-1.2092028,
 pem=-2.2721138, similarity=-19.639204|
 RECALL MISS: 296636 Power_outage]
 9: 37335 Drought
 [winner: 3672473 2006_Horn_of_Africa_food_crisis,
 pem_pos=3, final_scores=-1.2791578,
 final_local_scores=-1.261369, pem=-1.096551,
 similarity=-27.588842|
 gt: 37335 Drought, pem_pos=0, final_scores=-1.3781846,
 final_local_scores=-1.414727, pem=-0.049466744,
 similarity=-37.85956]
 10: 3292675 Reservoir
 [winner: 14524817 Jacqueline_Kennedy_Onassis_Reservoir,
 pem_pos=0, final_scores=-1.1347865,
 final_local_scores=-1.0705365, pem=0.0,
 similarity=-27.857231|
 RECALL MISS: 3292675 Reservoir]
 11: 381399 Hydroelectricity
 [winner: 14073 Hydropower, pem_pos=0,
 final_scores=-1.4039603, final_local_scores=-1.4449656,
 pem=0.0, similarity=-39.03094|
 gt: 381399 Hydroelectricity, pem_pos=1,
 final_scores=-1.536301, final_local_scores=-1.6063895,
 pem=-0.021431543, similarity=-43.731606]
 1: 322581 Phan_Văn_Khải
 [pem_pos=0, final_scores=0.53389466,
 final_local_scores=0.5729441, pem=0.0, similarity=21.187508]
 1: 202354 Vietnam
 [pem_pos=0, final_scores=0.46168607,
 final_local_scores=0.44229013, pem=-0.39026603,
 similarity=19.41086]
 2: 322581 Phan_Văn_Khải
 [pem_pos=0, final_scores=0.097849056,
 final_local_scores=0.09663051, pem=0.0, similarity=6.97336]
 3: 202354 Vietnam
 [pem_pos=0, final_scores=0.4883172,
 final_local_scores=0.4713806, pem=-0.39026603,
 similarity=20.278978]
 4: 3621429 Lê_Đức_Anh
 [pem_pos=0, final_scores=0.44431722,
 final_local_scores=0.50346404, pem=0.0, similarity=19.114084]
 5: 202354 Vietnam
 [pem_pos=0, final_scores=0.49406952,

final_local_scores=0.4776641, pem=-0.39026603,
similarity=20.46649]

Sample: b'APW19990519_0141_htm'

1 Astronomers Spot Storm on Mars [1 WASHINGTON 1] -LRB- AP -
RRB- -- A 2 cyclone four times the size of [1 1 Texas 1]
raged across the northern polar region of the planet [2 2 Mars
2] last month , according to 3 space telescope views of the
red planet . The [3 3 Space Telescope Science Institute 3]
announced today that the [4 4 Hubble Space Telescope 4] on
April 27 spotted an immense 4 Martian storm cloud , made up
of water 5 ice , that was 1,100 miles long and 900 wide .
Bands of clouds spiraled in a counterclockwise motion around
a 200-mile wide eye in the center of the storm , taking on a
shape that resembled 6 hurricanes on [5 5 Earth 5] ,
astronomers said . The storm occurred during the summer in [2
Mars 2] ' northern hemisphere . It came after seasonal
warming 7 evaporated the 8 carbon dioxide 9 ice sheet that
caps the 10 Martian north pole during part of the year .
Officials said the storm was three times larger than a storm
spotted 20 years ago by the [6 6 Viking Orbiter 6]
spacecraft . Astronomers used a 11 wide field planetary camera
on [3 Hubble 3] to photograph [4 Mars 4] during the planet 's
closest approach to [5 Earth 5] in nearly eight years . Two
pictures , taken six hours apart on April 27 , showed that
the storm moved slightly eastward . In the second picture ,
the storm appeared to be in the process of breaking up .
Later pictures by the [6 Hubble 6] camera failed to find the
storm , and astronomers said it could have been a short-lived
12 phenomenon .

1: 29810 Texas

2: 14640471 Mars

3: 177098 Space_Telescope_Science_Institute

4: 40203 Hubble_Space_Telescope

5: 9228 Earth

6: 38082 Viking_program

1: 580 Astronomer

[winner: 580 Astronomer, pem_pos=0,
final_scores=-0.94401187, final_local_scores=-1.0570942,
pem=-0.08467682, similarity=-26.995602]

gt: 580 Astronomer, pem_pos=0, final_scores=-0.94401187,
final_local_scores=-1.0570942, pem=-0.08467682,
similarity=-26.995602]

2: 42806 Cyclone

[winner: 21651467 Cyclone_(1987_film), pem_pos=13,
final_scores=-1.2901998, final_local_scores=-1.2634552,
pem=-3.269016, similarity=-15.836892]

gt: 42806 Cyclone, pem_pos=0, final_scores=-1.4180121,
final_local_scores=-1.513767, pem=-1.1976715,
similarity=-34.57099]

3: 29006 Space_observatory

[winner: 628467 Space_burial, pem_pos=0,

```

final_scores=-0.83026487, final_local_scores=-0.86618924,
pem=0.0, similarity=-21.759102|
    RECALL MISS: 29006 Space_observatory]
    [winner: 758427 Anglo-Australian_Telescope, pem_pos=3,
final_scores=-0.9798948, final_local_scores=-1.045064,
pem=-1.0981913, similarity=-21.124949|
    gt: 29006 Space_observatory, pem_pos=17,
final_scores=-2.3918648, final_local_scores=-2.6480758,
pem=-8.871224, similarity=-26.691057 ]
4: 47792 Martian
    [winner: 14640471 Mars, pem_pos=1,
final_scores=-0.047515616, final_local_scores=-0.15145738,
pem=-1.2732651, similarity=6.4941363|
    gt: 47792 Martian, pem_pos=0, final_scores=-0.10945226,
final_local_scores=-0.10579319, pem=-0.5593666,
similarity=3.9745553 ]
5: 14946 Ice
    [winner: 1889217 Italian_ice, pem_pos=0,
final_scores=-0.64557534, final_local_scores=-0.5601336,
pem=0.0, similarity=-12.625792|
    gt: 14946 Ice, pem_pos=1, final_scores=-0.87796444,
final_local_scores=-0.9656311, pem=-0.30294955,
similarity=-23.079159 ]
6: 8282374 Tropical_cyclone
    [winner: 8282374 Tropical_cyclone, pem_pos=0,
final_scores=-1.2460427, final_local_scores=-1.3222551,
pem=0.0, similarity=-35.369015|
    gt: 8282374 Tropical_cyclone, pem_pos=0,
final_scores=-1.2460427, final_local_scores=-1.3222551,
pem=0.0, similarity=-35.369015 ]
7: 10303 Evaporation
    [winner: 10303 Evaporation, pem_pos=0,
final_scores=-1.0535519, final_local_scores=-1.1021595,
pem=-0.037139524, similarity=-28.598955|
    gt: 10303 Evaporation, pem_pos=0,
final_scores=-1.0535519, final_local_scores=-1.1021595,
pem=-0.037139524, similarity=-28.598955 ]
8: 5906 Carbon_dioxide
    [winner: 59156 Dry_ice, pem_pos=0, final_scores=-0.73882,
final_local_scores=-0.76094675, pem=0.0,
similarity=-18.618456|
    RECALL MISS: 5906 Carbon_dioxide]
9: 493760 Ice_sheet
    [winner: 19745761 Sheet,_Hampshire, pem_pos=0,
final_scores=-0.644097, final_local_scores=-0.5133223,
pem=-0.4631585, similarity=-8.710121|
    RECALL MISS: 493760 Ice_sheet]
    [winner: 59156 Dry_ice, pem_pos=0, final_scores=-0.73882,
final_local_scores=-0.76094675, pem=0.0,
similarity=-18.618456|
    RECALL MISS: 493760 Ice_sheet]
10: 7723558 Planum_Boreum

```

[winner: 14640471 Mars, pem_pos=1,
final_scores=-0.013671294, final_local_scores=-0.1144876,
pem=-1.2732651, similarity=7.597388]

RECALL MISS: 7723558 Planum_Boreum]

[winner: 28040648 North_Pole_Idaho, pem_pos=2,
final_scores=-0.8515379, final_local_scores=-0.72653085,
pem=-1.8903615, similarity=-7.3113365]

RECALL MISS: 7723558 Planum_Boreum]

11: 885509 Wide_Field_and_Planetary_Camera

[winner: 305126 Field_of_view, pem_pos=0,
final_scores=-0.8465505, final_local_scores=-0.91068333,
pem=0.0, similarity=-23.086895]

RECALL MISS: 885509 Wide_Field_and_Planetary_Camera]

[winner: 385089 Planetary_(comics), pem_pos=0,
final_scores=-1.0630298, final_local_scores=-1.0681479,
pem=0.0, similarity=-27.78595]

RECALL MISS: 885509 Wide_Field_and_Planetary_Camera]

[winner: 52648 Camera, pem_pos=0,
final_scores=-1.1105411, final_local_scores=-1.1902447,
pem=-0.1663076, similarity=-30.525158]

RECALL MISS: 885509 Wide_Field_and_Planetary_Camera]

12: 60931 Phenomenon

[winner: 60931 Phenomenon, pem_pos=0,
final_scores=-1.2261351, final_local_scores=-1.3432693,
pem=-0.14534569, similarity=-35.205708]

gt: 60931 Phenomenon, pem_pos=0,
final_scores=-1.2261351, final_local_scores=-1.3432693,
pem=-0.14534569, similarity=-35.205708]

1: 29810 Texas

[pem_pos=0, final_scores=0.4750356,
final_local_scores=0.59293234, pem=-0.16191302,
similarity=22.664505]

2: 14640471 Mars

[pem_pos=0, final_scores=0.19654311,
final_local_scores=0.121477485, pem=-0.39377248,
similarity=9.856239]

3: 177098 Space_Telescope_Science_Institute

[pem_pos=0, final_scores=0.15419771,
final_local_scores=0.17590511, pem=0.0, similarity=9.339072]

4: 40203 Hubble_Space_Telescope

[pem_pos=0, final_scores=0.2555256,
final_local_scores=0.18987823, pem=0.0, similarity=9.756058]

5: 9228 Earth

[pem_pos=0, final_scores=0.12524687,
final_local_scores=0.060708597, pem=-0.07783105,
similarity=6.324636]

6: 38082 Viking_program

[pem_pos=0, final_scores=0.4054125,
final_local_scores=0.39070427, pem=0.0, similarity=15.749108]

1: 108956 Washington,_D.C.

[pem_pos=0, final_scores=0.08221854,
final_local_scores=0.15635467, pem=-1.7496481,

```

similarity=18.270506]
2: 14640471 Mars
    [pem_pos=0, final_scores=0.15649302,
final_local_scores=0.07772884, pem=-0.39377248,
similarity=8.550693]
3: 40203 Hubble_Space_Telescope
    [pem_pos=0, final_scores=0.15321393,
final_local_scores=0.078118294, pem=-0.20751034,
similarity=7.549392]
4: 14640471 Mars
    [pem_pos=0, final_scores=0.1218559,
final_local_scores=0.03989303, pem=-0.39377248,
similarity=7.4215965]
5: 9228 Earth
    [pem_pos=0, final_scores=0.16749571,
final_local_scores=0.10685906, pem=-0.07783105,
similarity=7.7018576]
6: 40203 Hubble_Space_Telescope
    [pem_pos=0, final_scores=0.16319151,
final_local_scores=0.08901727, pem=-0.20751034,
similarity=7.8746395]

```

Sample: b'APW19980625_1136_htm'

Report : Plane robbed by gunmen in [1 Nigerian 1] airport An [1 1 Air Afrique 1] [2 2 Boeing 2] -727 jet was the third passenger liner looted in the past month by armed robbers while awaiting takeoff at [2 Nigeria 2] 's largest international airport , the 1 Lagos 2 Guardian newspaper reported on Thursday . The thieves broke into the aircraft 's luggage compartment and escaped with a large quantity of baggage as the plane was awaiting permission to take off from the [3 Lagos 3] international airport control tower around 9 p.m. on Tuesday night , the [4 Guardian 4] said . It was not clear where the airliner was destined . There was no immediate confirmation of the report . The plane 's captain , who noticed the theft , sent a 3 distress call that went unanswered by airport officials . 4 Security personnel on duty later said said they were unable to capture the culprits because their only security vehicle had been taken away for routine maintenance . In another incident last week involving an unidentified passenger jet , landing gear worth dlrs 1 million was stripped from while the plane was waiting to take off , the newspaper said . In May , a [3 3 Cameroon Airlines 3] plane was robbed of its luggage after armed thieves blocked the plane on the 5 tarmac with a large log . The robberies came despite orders from airport security officials to shoot suspected thieves on sight . [4 4 Nigeria 4] , [5 Africa 5] 's most populous country , has a notoriously high rate of 6 violent crime accompanied by a breakdown of law and order under years of military rule . -LRB- fa-gam -RRB-

```

1: 2108164 Air_Afrique
2: 18933266 Boeing

```

3: 1459639 Cameroon_Airlines

4: 21383 Nigeria

1: 85232 Lagos

[winner: 85232 Lagos, pem_pos=0, final_scores=-1.0379303, final_local_scores=-1.2040436, pem=-0.20534219, similarity=-30.72467|

gt: 85232 Lagos, pem_pos=0, final_scores=-1.0379303, final_local_scores=-1.2040436, pem=-0.20534219, similarity=-30.72467]

2: 7397412 The_Guardian_(Nigeria)

[winner: 19344515 The_Guardian, pem_pos=0, final_scores=-1.1235406, final_local_scores=-1.1954432, pem=-0.6935336, similarity=-27.813154|

gt: 7397412 The_Guardian_(Nigeria), pem_pos=8, final_scores=-1.4655759, final_local_scores=-1.5369864, pem=-3.581484, similarity=-22.300354]

[winner: 27830504 Drita_(newspaper), pem_pos=13, final_scores=-1.2416053, final_local_scores=-1.1369057, pem=-3.1345718, similarity=-12.791529|

RECALL MISS: 7397412 The_Guardian_(Nigeria)]

3: 303507 Distress_signal

[winner: 14230294 Distress_(medicine), pem_pos=1, final_scores=-1.2852993, final_local_scores=-1.2131071, pem=-1.0111346, similarity=-26.61312|

gt: 303507 Distress_signal, pem_pos=0, final_scores=-1.311249, final_local_scores=-1.331506, pem=0.0, similarity=-35.645084]

[winner: 204251 Call_option, pem_pos=1, final_scores=-1.4101624, final_local_scores=-1.4177034, pem=-0.96357596, similarity=-32.977306|

RECALL MISS: 303507 Distress_signal]

4: 284770 Airport_security

[winner: 41684 Security, pem_pos=0, final_scores=-0.42215362, final_local_scores=-0.41904414, pem=-1.0129223, similarity=-2.9069717|

RECALL MISS: 284770 Airport_security]

5: 167921 Tarmac

[winner: 167921 Tarmac, pem_pos=0, final_scores=-1.3519909, final_local_scores=-1.3988898, pem=-0.024466075, similarity=-37.5229|

gt: 167921 Tarmac, pem_pos=0, final_scores=-1.3519909, final_local_scores=-1.3988898, pem=-0.024466075, similarity=-37.5229]

6: 273381 Violent_crime

[winner: 4035351 Birmingham_campaign, pem_pos=2, final_scores=-1.0730653, final_local_scores=-1.0892863, pem=-0.691789, similarity=-24.654707|

gt: 273381 Violent_crime, pem_pos=4, final_scores=-1.687279, final_local_scores=-1.7416804, pem=-3.5098479, similarity=-28.798403]

[winner: 26060819 Crime_in_Kansas, pem_pos=7, final_scores=-1.1577951, final_local_scores=-1.0603368,

pem=-2.8839502, similarity=-11.869483|
 RECALL MISS: 273381 Violent_crime]

1: 2108164 Air_Afrique
 [pem_pos=0, final_scores=0.036609635,
 final_local_scores=0.070736915, pem=0.0, similarity=6.2006435]

2: 18933266 Boeing
 [pem_pos=0, final_scores=0.061363444,
 final_local_scores=0.11190194, pem=0.0, similarity=7.42909]

3: 1459639 Cameroon_Airlines
 [pem_pos=0, final_scores=0.27619308,
 final_local_scores=0.33496195, pem=0.0, similarity=14.085646]

4: 21383 Nigeria
 [pem_pos=0, final_scores=0.54517895,
 final_local_scores=0.5202442, pem=-0.31455088,
 similarity=21.325415]

1: 21383 Nigeria
 [pem_pos=0, final_scores=0.5707585,
 final_local_scores=0.548186, pem=-0.075027294,
 similarity=20.856688]

2: 21383 Nigeria
 [pem_pos=0, final_scores=0.5120649,
 final_local_scores=0.48407215, pem=-0.31455088,
 similarity=20.245968]

3: 85232 Lagos
 [pem_pos=0, final_scores=0.33489662,
 final_local_scores=0.3140904, pem=-0.20534219,
 similarity=14.57948]

4: 19344515 The_Guardian
 [pem_pos=0, final_scores=0.1939433,
 final_local_scores=0.2850927, pem=-0.6935336,
 similarity=16.368992]

5: 5334607 Africa
 [pem_pos=0, final_scores=0.5225048,
 final_local_scores=0.5317777, pem=-0.13197383,
 similarity=20.676714]

Sample: b'APW19981026_0096_htm'

Report : [1 1 South Africa 1] n truth panel to condemn [1
 ANC 1] A panel probing apartheid - era abuses has accused the
 [2 2 African National Congress 2] of 1 human rights
 violations , including 2 torture and 3 bomb attacks , the
 state 4 broadcaster said Monday . The [2 ANC 2] , which led
 the struggle against 5 white rule and now is in power ,
 previously has acknowledged it was told to expect implication
 in right violations . The 6 Truth and Reconciliation Commissi
 on will release its final report on 2 1-2 years of
 investigation on Thursday . The report was expected to place
 overwhelming blame for the era 's abuses on white governments
 and their brutal security services . 7 South African Broadcas
 ting Corp. . radio said it had been leaked a `` preliminary
 document '' which condemns the [3 ANC 3] as politically and
 morally responsible for gross human rights violations during

and after the fall of apartheid . An [4 ANC 4] spokesman , Thabo Masebe , said he believed the news report referred to a commission letter sent in advance `` informing us of its intention to implicate the [5 ANC 5] in gross human rights violations . '' [6 Commission 6] officials could not immediately be reached for comment . The [7 ANC 7] was nervous enough about the report to request a meeting with 8 commissioners earlier this month . The panel refused , to avoid appearing embarrassed . The party says the accusations against it likely would center on the planting of 9 land mines on border farms , abuses at its military camps in [3 3 Angola 3] and bombings . -LRB- djw -RRB-

1: 17416221 South_Africa

2: 2503 African_National_Congress

3: 701 Angola

1: 13831 Human_rights

[winner: 51490 Rights, pem_pos=0, final_scores=-1.1836586, final_local_scores=-1.2270529, pem=-0.3661264, similarity=-30.536942]

gt: 13831 Human_rights, pem_pos=1, final_scores=-1.592927, final_local_scores=-1.7362642, pem=-2.168652, similarity=-35.93041]

[winner: 4397438 Human_(Dungeons_&_Dragons), pem_pos=2, final_scores=-1.3513161, final_local_scores=-1.320478, pem=-1.9270188, similarity=-24.836555]

gt: 13831 Human_rights, pem_pos=1, final_scores=-1.6728888, final_local_scores=-1.8236103, pem=-1.8663583, similarity=-40.18091]

2: 47702 Torture

[winner: 10058543 Human_rights_in_Zimbabwe, pem_pos=3, final_scores=-1.0744827, final_local_scores=-1.1518092, pem=-1.3860477, similarity=-22.745026]

gt: 47702 Torture, pem_pos=0, final_scores=-1.1473209, final_local_scores=-1.2062528, pem=-0.069790564, similarity=-31.527744]

3: 47628 Bomb

[winner: 47628 Bomb, pem_pos=0, final_scores=-1.2835782, final_local_scores=-1.3292937, pem=-0.20826642, similarity=-34.44648]

gt: 47628 Bomb, pem_pos=0, final_scores=-1.2835782, final_local_scores=-1.3292937, pem=-0.20826642, similarity=-34.44648]

4: 113604 Broadcasting

[winner: 521153 Ford_C._Frick_Award, pem_pos=0, final_scores=-1.3668014, final_local_scores=-1.3704326, pem=0.0, similarity=-36.80673]

gt: 113604 Broadcasting, pem_pos=2, final_scores=-1.826518, final_local_scores=-1.9110024, pem=-1.3794667, similarity=-45.436653]

5: 888454 Dominant_minority

[winner: 888454 Dominant_minority, pem_pos=0, final_scores=-0.715302, final_local_scores=-0.7655882,

pem=0.0, similarity=-18.756968|
gt: 888454 Dominant_minority, pem_pos=0,
final_scores=-0.715302, final_local_scores=-0.7655882,
pem=0.0, similarity=-18.756968]
6: 30865 Truth_and_Reconciliation_Commission_(South_Africa)
[winner: 3125362 Reconciliation_and_Unity_Commission,
pem_pos=0, final_scores=-0.375769,
final_local_scores=-0.3470626, pem=-0.18232158,
similarity=-5.2758336|
gt: 30865
Truth_and_Reconciliation_Commission_(South_Africa),
pem_pos=1, final_scores=-0.5321471,
final_local_scores=-0.5917336, pem=-1.7917595,
similarity=-3.82493]
[winner: 29920 Truth, pem_pos=0,
final_scores=-0.97184455, final_local_scores=-0.9329241,
pem=-0.3064552, similarity=-22.084051|
RECALL MISS: 30865
Truth_and_Reconciliation_Commission_(South_Africa)]
7: 364060 South_African_Broadcasting_Corporation
[winner: 364060 South_African_Broadcasting_Corporation,
pem_pos=0, final_scores=-1.0574751,
final_local_scores=-1.1682804, pem=0.0,
similarity=-30.774103|
gt: 364060 South_African_Broadcasting_Corporation,
pem_pos=0, final_scores=-1.0574751,
final_local_scores=-1.1682804, pem=0.0,
similarity=-30.774103]
[winner: 7485 Corporation, pem_pos=0,
final_scores=-1.1335688, final_local_scores=-1.1923828,
pem=-0.005830923, similarity=-31.46166|
RECALL MISS: 364060
South_African_Broadcasting_Corporation]
[winner: 4721653 Zyphe_language, pem_pos=12,
final_scores=-1.4250015, final_local_scores=-1.3942618,
pem=-4.4188404, similarity=-13.487497|
RECALL MISS: 364060
South_African_Broadcasting_Corporation]
8: 431272 Commissioner
[winner: 23385143
High_Commissioners_for_Palestine_and_Transjordan, pem_pos=5,
final_scores=-1.2157031, final_local_scores=-1.2136707,
pem=-1.609438, similarity=-23.376268|
gt: 431272 Commissioner, pem_pos=1,
final_scores=-1.6663393, final_local_scores=-1.7585604,
pem=-1.3210635, similarity=-41.205086]
9: 18172 Land_mine
[winner: 989207 Anti-tank_mine, pem_pos=2,
final_scores=-0.83564794, final_local_scores=-0.7976153,
pem=-1.0893145, similarity=-13.788857|
gt: 18172 Land_mine, pem_pos=4, final_scores=-1.1413736,
final_local_scores=-1.1884096, pem=-2.4937956,

```

similarity=-17.813152 ]
[winner: 7948840 Bhūmi, pem_pos=10,
final_scores=-1.313324, final_local_scores=-1.234247,
pem=-2.4827268, similarity=-19.241226]
RECALL MISS: 18172 Land_mine]
1: 17416221 South_Africa
[pem_pos=0, final_scores=0.39935535,
final_local_scores=0.42467332, pem=-0.26251382,
similarity=18.190403]
2: 2503 African_National_Congress
[pem_pos=0, final_scores=0.46293324,
final_local_scores=0.38923943, pem=0.0, similarity=15.705394]
3: 701 Angola
[pem_pos=0, final_scores=0.48173875,
final_local_scores=0.53765655, pem=-0.39249876,
similarity=22.268927]
1: 2503 African_National_Congress
[pem_pos=0, final_scores=0.46883124,
final_local_scores=0.39568204, pem=-0.05902751,
similarity=16.218655]
2: 2503 African_National_Congress
[pem_pos=0, final_scores=0.50079465,
final_local_scores=0.43059725, pem=-0.05902751,
similarity=17.260593]
3: 2503 African_National_Congress
[pem_pos=0, final_scores=0.50066596,
final_local_scores=0.4304567, pem=-0.05902751,
similarity=17.2564]
4: 2503 African_National_Congress
[pem_pos=0, final_scores=0.4716192,
final_local_scores=0.39872748, pem=-0.05902751,
similarity=16.309536]
5: 2503 African_National_Congress
[pem_pos=0, final_scores=0.48418128,
final_local_scores=0.41244966, pem=-0.05902751,
similarity=16.719034]
6: 9974 European_Commission
[pem_pos=0, final_scores=0.343059,
final_local_scores=0.42317992, pem=-0.76247066,
similarity=20.864681]
7: 2503 African_National_Congress
[pem_pos=0, final_scores=0.47025073,
final_local_scores=0.39723265, pem=-0.05902751,
similarity=16.264929]

```

Sample: b'APW19981109_1089_htm'

[1 LONDON 1] : Nov. 26 . The apparent settlement came after four days of talks involving the [2 1 West Indies 2] players and officials of the [3 West Indies 3] and [4 2 South Africa 4] 3 cricket boards . The crisis came to a head last Wednesday when the [5 West Indies Cricket Board 5] fired superstar 4 batsman [1 1 Brian Lara 1] as captain and [2 2 Carl Hooper 2]

as vice-captain . The two had refused to travel to [6 South Africa 6] , demanding better pay and conditions for the tour . Team members have insisted that [7 Lara 7] and [8 Hooper 8] be reinstated as part of any settlement . [3 3 Jimmy Adams 3] , treasurer of the players ' association , negotiated on behalf of the players Monday with Pat Rousseau , president of the [4 4 West Indies Cricket Board 4] . [5 5 Ali Bacher 5] , head of [9 5 South African 9] cricket , also was directly involved in Monday 's talks after sitting on the sidelines Sunday . [10 West Indies 10] tour manager [6 6 Clive Lloyd 6] and [11 WICB 11] member [7 7 Joel Garner 7] have also been involved in the talks which began Friday . Progress was reported over the weekend with the announcement of a 6 sponso rship deal offered by former [12 South African 12] 7 wicketke eper David Richardson , who runs a 8 sports marketing firm in [8 8 Johannesburg 8] . Earnings from the sponsorship proposal would apparently help meet the players ' demands for better pay , which is reported to range from dlrs 15,000 to 60,000 for the [13 South African 13] tour . -LRB- jp/sw -RRB-

1: 4367 Brian_Lara

2: 2112133 Carl_Hooper

3: 1026319 Jimmy_Adams

4: 8381696 West_Indies_Cricket_Board

5: 3034646 Ali_Bacher

6: 624962 Clive_Lloyd

7: 87055 Joel_Garner

8: 58893 Johannesburg

1: 18956035 Caribbean

[winner: 3379941 West_Indies_cricket_team, pem_pos=1, final_scores=0.292906, final_local_scores=0.2266767, pem=-1.1020415, similarity=16.847271|

gt: 18956035 Caribbean, pem_pos=0, final_scores=-0.24077187, final_local_scores=-0.23826039, pem=-0.63651145, similarity=0.44099596]

2: 17416221 South_Africa

[winner: 759263 South_Africa_national_cricket_team, pem_pos=1, final_scores=0.25695854, final_local_scores=0.1951431, pem=-3.0105038, similarity=26.28476|

gt: 17416221 South_Africa, pem_pos=0, final_scores=-0.08217023, final_local_scores=-0.099929966, pem=-0.26251382, similarity=2.5351944]

3: 25675557 Cricket

[winner: 330091 Rhaphidophoridae, pem_pos=10, final_scores=-1.4567815, final_local_scores=-1.4334683, pem=-2.56456, similarity=-24.741367|

gt: 25675557 Cricket, pem_pos=0, final_scores=-1.8946345, final_local_scores=-2.1121337, pem=-0.07532314, similarity=-58.530952]

4: 673526 Batting_(cricket)

[winner: 673526 Batting_(cricket), pem_pos=0, final_scores=-1.3849384, final_local_scores=-1.5187881,

pem=-0.0001237469, similarity=-41.233284|
gt: 673526 Batting_(cricket), pem_pos=0,
final_scores=-1.3849384, final_local_scores=-1.5187881,
pem=-0.0001237469, similarity=-41.233284]
5: 759263 South_Africa_national_cricket_team
[winner: 17416221 South_Africa, pem_pos=0,
final_scores=0.29534543, final_local_scores=0.32485655,
pem=-0.1389451, similarity=14.539685|
gt: 759263 South_Africa_national_cricket_team,
pem_pos=1, final_scores=-0.10478677,
final_local_scores=-0.20539278, pem=-2.617051,
similarity=12.192312]
[winner: 330091 Rhabdophoridae, pem_pos=10,
final_scores=-1.5483706, final_local_scores=-1.5335155,
pem=-2.56456, similarity=-27.726969|
RECALL MISS: 759263 South_Africa_national_cricket_team]
6: 545607 Sponsor_(commercial)
[winner: 29858442 Sports_marketing, pem_pos=1,
final_scores=-1.2971687, final_local_scores=-1.309237,
pem=-0.69121295, similarity=-31.221607|
gt: 545607 Sponsor_(commercial), pem_pos=0,
final_scores=-1.4466, final_local_scores=-1.5015824, pem=0.0,
similarity=-40.720497]
7: 463313 Wicket-keeper
[winner: 463313 Wicket-keeper, pem_pos=0,
final_scores=-1.7486626, final_local_scores=-1.9508271,
pem=0.0, similarity=-54.12686|
gt: 463313 Wicket-keeper, pem_pos=0,
final_scores=-1.7486626, final_local_scores=-1.9508271,
pem=0.0, similarity=-54.12686]
8: 29858442 Sports_marketing
[winner: 29858442 Sports_marketing, pem_pos=0,
final_scores=-0.95937544, final_local_scores=-0.94024926,
pem=0.0, similarity=-23.9692|
gt: 29858442 Sports_marketing, pem_pos=0,
final_scores=-0.95937544, final_local_scores=-0.94024926,
pem=0.0, similarity=-23.9692]
1: 4367 Brian_Lara
[pem_pos=0, final_scores=0.8286579,
final_local_scores=0.83703285, pem=0.0, similarity=29.068443]
2: 2112133 Carl_Hooper
[pem_pos=0, final_scores=0.71943337,
final_local_scores=0.74795175, pem=0.0, similarity=26.410086]
3: 1026319 Jimmy_Adams
[pem_pos=0, final_scores=0.43801087,
final_local_scores=0.4501438, pem=0.0, similarity=17.522902]
4: 8381696 West_Indies_Cricket_Board
[pem_pos=0, final_scores=0.24860372,
final_local_scores=0.22938764, pem=0.0, similarity=10.935098]
5: 3034646 Ali_Bacher
[pem_pos=0, final_scores=0.7128177,
final_local_scores=0.7625129, pem=0.0, similarity=26.84462]

6: 624962 Clive_Lloyd
[pem_pos=0, final_scores=0.79719406,
final_local_scores=0.8231863, pem=0.0, similarity=28.655233]
7: 87055 Joel_Garner
[pem_pos=0, final_scores=0.72963333,
final_local_scores=0.774887, pem=0.0, similarity=27.213888]
8: 58893 Johannesburg
[pem_pos=0, final_scores=0.43733376,
final_local_scores=0.47040766, pem=-0.09562059,
similarity=18.647615]
1: 17867 London
[pem_pos=0, final_scores=0.5328278,
final_local_scores=0.6003152, pem=-0.09086737,
similarity=22.498466]
2: 3379941 West_Indies_cricket_team
[pem_pos=1, final_scores=0.292906,
final_local_scores=0.2266767, pem=-1.1020415,
similarity=16.847271]
3: 3379941 West_Indies_cricket_team
[pem_pos=1, final_scores=0.101869985,
final_local_scores=0.017998904, pem=-1.1020415,
similarity=10.619909]
4: 759263 South_Africa_national_cricket_team
[pem_pos=1, final_scores=0.25695854,
final_local_scores=0.1951431, pem=-3.0105038,
similarity=26.28476]
5: 8381696 West_Indies_Cricket_Board
[pem_pos=0, final_scores=0.31547332,
final_local_scores=0.3024325, pem=0.0, similarity=13.1149025]
6: 17416221 South_Africa
[pem_pos=0, final_scores=0.44046497,
final_local_scores=0.48337764, pem=-0.26251382,
similarity=19.942257]
7: 4367 Brian_Lara
[pem_pos=0, final_scores=0.7764283,
final_local_scores=0.78622025, pem=0.0, similarity=27.552094]
8: 2112133 Carl_Hooper
[pem_pos=0, final_scores=0.39772946,
final_local_scores=0.39653903, pem=0.0, similarity=15.923228]
9: 17416221 South_Africa
[pem_pos=0, final_scores=0.29534543,
final_local_scores=0.32485655, pem=-0.1389451,
similarity=14.539685]
10: 3379941 West_Indies_cricket_team
[pem_pos=1, final_scores=0.40020496,
final_local_scores=0.34388453, pem=-1.1020415,
similarity=20.344986]
11: 3379941 West_Indies_cricket_team
[pem_pos=2, final_scores=0.19868062,
final_local_scores=0.12355253, pem=-2.0165277,
similarity=18.742958]
12: 759263 South_Africa_national_cricket_team

[pem_pos=1, final_scores=0.086371645,
final_local_scores=0.0073881745, pem=-2.617051,
similarity=18.54212]
13: 17416221 South_Africa
[pem_pos=0, final_scores=0.27722853,
final_local_scores=0.3050666, pem=-0.1389451,
similarity=13.949114]

Sample: b'APW19980816_0994_htm'

Stolen painting in [1 1 New Zealand 1] recovered , one man arrested A 19th-century painting worth [1 U.S. 1] dlrs 1 million stolen from [2 2 Auckland Art Gallery 2] nine days ago has been recovered , police said Monday . Senior Sgt. Gary Allcock said he understood that one person had been arrested over the theft of the 1874 oil , `` Still on Top , '' by [2 French-born 2] artist 1 Jacques-Joseph -LRB- James -RRB- Tissot , one of [3 New Zealand 3] 's most expensive and popular [4 European 4] works . He said details would be disclosed later . The painting valued at up to [5 U.S. 5] dlrs 1 million was stolen Aug. 9 by a man armed with a 2 shot gun in a daring robbery at the [6 3 Auckland Art Gallery 6] . The man , clad in black and wearing a motorcycle helmet , dark visor and gloves , burst through the main door with a sawed-off shotgun and a 4 crowbar . Museum official Kate Darrow said the man ran through the Grey Gallery waving a 5 firearm and yelling `` get down , '' and `` keep back '' to staff and patrons . She said the man appeared to know where he was going and went to the Lower Wellesley Gallery , where the painting was displayed . The gunman grabbed the painting off the wall and dropped it on the ground . He then used the crowbar to remove the canvas from the frame and take it off the backing board . The robber left through the main entrance and skirted behind the gallery into [7 6 Albert Park 7] , pursued by a member of the public . He was believed to have climbed on a motorcycle and escaped via a nearby highway .

1: 4913064 New_Zealand

2: 3660513 Auckland_Art_Gallery

1: 1440880 James_Tissot

[winner: 2549242 Jacques_Joseph_Champollion-Figeac,
pem_pos=0, final_scores=-0.51226383,
final_local_scores=-0.44112176, pem=0.0, similarity=-9.07424|
RECALL MISS: 1440880 James_Tissot]

[winner: 754118 Tissot, pem_pos=0,
final_scores=-0.5151837, final_local_scores=-0.45572913,
pem=0.0, similarity=-9.510153|

gt: 1440880 James_Tissot, pem_pos=1,
final_scores=-0.7288384, final_local_scores=-0.7152037,
pem=-1.4637004, similarity=-9.293563]

[winner: 224020 James_River, pem_pos=5,
final_scores=-1.0424325, final_local_scores=-1.0892103,
pem=-3.2486143, similarity=-10.748027|

RECALL MISS: 1440880 James_Tissot]

2: 26838 Shotgun
[winner: 26838 Shotgun, pem_pos=0,
final_scores=-1.050404, final_local_scores=-1.049494,
pem=-0.08884865, similarity=-26.746107|
gt: 26838 Shotgun, pem_pos=0, final_scores=-1.050404,
final_local_scores=-1.049494, pem=-0.08884865,
similarity=-26.746107]
3: 18660332 Auckland
[winner: 3660513 Auckland_Art_Gallery, pem_pos=0,
final_scores=0.24344288, final_local_scores=0.30235505,
pem=0.0, similarity=13.112591|
RECALL MISS: 18660332 Auckland]
4: 1950953 Crowbar_(tool)
[winner: 8257469 Spud_bar, pem_pos=7,
final_scores=-1.0510371, final_local_scores=-0.9668578,
pem=-2.1972246, similarity=-12.814404|
gt: 1950953 Crowbar_(tool), pem_pos=1,
final_scores=-1.1399559, final_local_scores=-1.0739677,
pem=-1.1161054, similarity=-21.89007]
5: 11966 Firearm
[winner: 11966 Firearm, pem_pos=0,
final_scores=-1.0152338, final_local_scores=-1.0501146,
pem=-0.002285169, similarity=-27.235376|
gt: 11966 Firearm, pem_pos=0, final_scores=-1.0152338,
final_local_scores=-1.0501146, pem=-0.002285169,
similarity=-27.235376]
6: 2986192 Albert_Park,_Auckland
[winner: 510140 Albert_Park,_Victoria, pem_pos=0,
final_scores=0.09472664, final_local_scores=0.13863072,
pem=-1.0991244, similarity=14.203939|
gt: 2986192 Albert_Park,_Auckland, pem_pos=4,
final_scores=-0.11331214, final_local_scores=-0.05196649,
pem=-1.9415679, similarity=13.097477]
1: 4913064 New_Zealand
[pem_pos=0, final_scores=0.5163131,
final_local_scores=0.49736017, pem=-0.23795462,
similarity=20.225967]
2: 3660513 Auckland_Art_Gallery
[pem_pos=0, final_scores=0.16899393,
final_local_scores=0.22103089, pem=0.0, similarity=10.685716]
1: 3434750 United_States
[pem_pos=0, final_scores=0.3789606,
final_local_scores=0.35217473, pem=-0.14075686,
similarity=15.364766]
2: 5843419 France
[pem_pos=0, final_scores=0.2736534,
final_local_scores=0.26716647, pem=0.0, similarity=12.062493]
3: 4913064 New_Zealand
[pem_pos=0, final_scores=0.5403809,
final_local_scores=0.5236506, pem=-0.23795462,
similarity=21.010525]
4: 9239 Europe

[pem_pos=0, final_scores=0.51370126,
final_local_scores=0.52709526, pem=-0.25557294,
similarity=21.209131]
5: 3434750 United_States
[pem_pos=0, final_scores=0.3866585,
final_local_scores=0.36058345, pem=-0.14075686,
similarity=15.6157]
6: 3660513 Auckland_Art_Gallery
[pem_pos=0, final_scores=0.24344288,
final_local_scores=0.30235505, pem=0.0, similarity=13.112591]
7: 510140 Albert_Park,_Victoria
[pem_pos=0, final_scores=0.09472664,
final_local_scores=0.13863072, pem=-1.0991244,
similarity=14.203939]

Sample: b'APW19980615_0417_hm'

Actress [1 Brooke Shields 1] accepts 1 libel 2 damages ,
apology over false Actress [1 1 Brooke Shields 1] on Monday
accepted `` very substantial '' libel damages and a public
apology over a newspaper article alleging that she was
questioned by police for two hours in connection with 3 drug
offenses as she left the [2 2 Cannes Film Festival 2] . [2 4 T
he Mail 2] on Sunday reported May 24 in an `` exclusive ''
that began on the front page and continued on page three that
Miss Shields ' luggage was searched and her flight at the [3 3
Nice Airport 3] delayed while [3 5 French 3] police quizzed
her . After the actress filed suit against [4 Associated
Newspapers Ltd 4] . , the publisher of the [5 Mail on Sunday 5
] , the paper 's editor and two Mail journalists , the [6
Mail on Sunday 6] published a front-page apology and 6 retrac
tion on May 31 . The retraction said the story was totally
false and conceded Miss Shields had not even been stopped at
the airport . `` The defendants recognize that Miss Shields
is a staunch and committed anti-drug campaigner and that she
has spent considerable amounts of time supporting 7 substance
abuse programs and raising much needed money for anti-drug
abuse programs , '' said Claire Shields , the defendants '
attorney . The newspaper 's front-page article `` must
therefore have been especially hurtful to her and to her
family , '' the attorney said . The Sunday paper 's editor ,
Jonathan Holborow , has written the actress , who was not
present in court Monday , a personal letter of apology . The
amount of the damages was not disclosed . -LRB- kg -RRB-

1: 18984642 Brooke_Shields
2: 101460 Cannes_Film_Festival
3: 291627 Nice_Côte_d'Azur_Airport
1: 28661 Defamation

[winner: 28661 Defamation, pem_pos=0,
final_scores=-1.180818, final_local_scores=-1.2850056,
pem=-0.010153767, similarity=-34.202194]
gt: 28661 Defamation, pem_pos=0, final_scores=-1.180818,
final_local_scores=-1.2850056, pem=-0.010153767,

similarity=-34.202194]
2: 8134 Damages
[winner: 13915654 Damages_(Jewish_law), pem_pos=0,
final_scores=-0.94729525, final_local_scores=-0.8631693,
pem=0.0, similarity=-21.668982|
gt: 8134 Damages, pem_pos=1, final_scores=-1.1800272,
final_local_scores=-1.2190015, pem=-0.026280418,
similarity=-32.144806]
3: 9311172 Drug
[winner: 803443 ATC_code_G, pem_pos=10,
final_scores=-0.92366797, final_local_scores=-0.69210774,
pem=-3.014498, similarity=-0.17085394|
gt: 9311172 Drug, pem_pos=0, final_scores=-1.3194499,
final_local_scores=-1.340538, pem=-1.2112672,
similarity=-29.327557]
4: 8944898 The_Mail_on_Sunday
[winner: 177181 Daily_Mail, pem_pos=0,
final_scores=0.46641213, final_local_scores=0.41762865,
pem=0.0, similarity=16.552586|
RECALL MISS: 8944898 The_Mail_on_Sunday]
[winner: 908910 Sunday_shopping, pem_pos=0,
final_scores=-1.0150137, final_local_scores=-1.0354564,
pem=0.0, similarity=-26.810371|
RECALL MISS: 8944898 The_Mail_on_Sunday]
5: 614023 National_Police_(France)
[winner: 5843419 France, pem_pos=0,
final_scores=0.64197826, final_local_scores=0.7499632,
pem=-0.80223274, similarity=30.832779|
RECALL MISS: 614023 National_Police_(France)]
[winner: 23627 Police, pem_pos=0,
final_scores=-1.4930587, final_local_scores=-1.5985267,
pem=-0.116511196, similarity=-42.979908|
gt: 614023 National_Police_(France), pem_pos=8,
final_scores=-1.9641986, final_local_scores=-2.0354252,
pem=-4.0567083, similarity=-34.590424]
6: 582691 Retraction
[winner: 4967125 Retract_(group_theory), pem_pos=2,
final_scores=-0.9842153, final_local_scores=-0.85777426,
pem=-0.6931472, similarity=-17.73854|
gt: 582691 Retraction, pem_pos=0,
final_scores=-1.1902633, final_local_scores=-1.1746714,
pem=-0.3513979, similarity=-29.053871]
7: 102959 Substance_abuse
[winner: 27568 Substance_theory, pem_pos=0,
final_scores=-1.1350738, final_local_scores=-1.1018306,
pem=-0.029888304, similarity=-28.628572|
gt: 102959 Substance_abuse, pem_pos=5,
final_scores=-2.2197895, final_local_scores=-2.3725612,
pem=-4.643808, similarity=-41.458492]
[winner: 424739 Abuse, pem_pos=0,
final_scores=-1.3992248, final_local_scores=-1.4826536,
pem=-0.669883, similarity=-36.512703|

```

gt: 102959 Substance_abuse, pem_pos=2,
final_scores=-1.6267874, final_local_scores=-1.7247965,
pem=-1.2310537, similarity=-40.686996 ]
1: 18984642 Brooke_Shields
[pem_pos=0, final_scores=0.27576143,
final_local_scores=0.3010533, pem=0.0, similarity=13.073744]
2: 101460 Cannes_Film_Festival
[pem_pos=0, final_scores=0.15511839,
final_local_scores=0.16329256, pem=-0.046005446,
similarity=9.212873]
3: 291627 Nice_Côte_d'Azur_Airport
[pem_pos=0, final_scores=0.046790227,
final_local_scores=0.18304023, pem=0.0, similarity=9.551998]
1: 18984642 Brooke_Shields
[pem_pos=0, final_scores=0.2784691,
final_local_scores=0.304011, pem=0.0, similarity=13.162007]
2: 177181 Daily_Mail
[pem_pos=0, final_scores=0.46641213,
final_local_scores=0.41762865, pem=0.0, similarity=16.552586]
3: 5843419 France
[pem_pos=0, final_scores=0.64197826,
final_local_scores=0.7499632, pem=-0.80223274,
similarity=30.832779]
4: 714372 DMG_Media
[pem_pos=0, final_scores=0.1685199,
final_local_scores=0.18897599, pem=0.0, similarity=9.729134]
5: 8944898 The_Mail_on_Sunday
[pem_pos=0, final_scores=0.26883823,
final_local_scores=0.2164025, pem=0.0, similarity=10.547595]
6: 8944898 The_Mail_on_Sunday
[pem_pos=0, final_scores=0.29332113,
final_local_scores=0.24217328, pem=0.0, similarity=11.316647]

```

Sample: b'APW19980603_0791_htm'

[1 THAILAND 1] MORE and more people are dying or suffering from fatal diseases such as 1 brain tumours , mammarian or 2 breast cancer and brain 3 haemorrhage due to a lack of proper medical care. However , things may change as the Srisiam Hospital in [1 1 Bangkok 1] has now established the first ever 4 Radio Surgery Institute in [2 2 Thailand 2] -- -LRB- there are presently more than 100 radio surgery centres around the world -RRB- . Radio surgery is a modern technique which does n't require surgery in order to remove 5 cancerous growth or 6 tissues . It involves a 7 three dimensional method known as 8 Stereotaxic Surgery , which is also used in conjunction with 9 radiation equipment known as the 10 Linear Accelerator . With this new technology , doctors are able to removal cancerous tissues embedded deep inside the brain or near sensitive organs that makes normal surgery very difficult and risky . With radio surgery , every step is controlled by computers and patients are 11 conscious while they undergo treatment . This new technique is also capable

of treating diseases such as Artero Venous 12 Malformation - LRB- the 13 clotting of veins or blood 14 capillaries in the brain -RRB- . So far , ninety per cent of the radio surgery treatments have been successful and patients only have to remain in hospital for three days . Additionally , there is no need for patients to shave their heads and radio surgery causes no scar , little blood loss and is not expensive.For more information contact the hospital at srisiam ksc.th.com .

1: 56615 Bangkok

2: 30128 Thailand

1: 37284 Brain_tumor

[winner: 1236730 Neoplasm, pem_pos=0, final_scores=-1.1206025, final_local_scores=-1.1119273, pem=-0.005333375, similarity=-29.06341]

gt: 37284 Brain_tumor, pem_pos=3, final_scores=-2.5770092, final_local_scores=-2.7367268, pem=-6.6227365, similarity=-41.5642]

[winner: 3717 Brain, pem_pos=1, final_scores=-1.120954, final_local_scores=-1.1352488, pem=-0.1751658, similarity=-28.835796]

gt: 37284 Brain_tumor, pem_pos=0, final_scores=-1.3290237, final_local_scores=-1.3734922, pem=0.0, similarity=-36.898037]

2: 70547 Breast_cancer

[winner: 4489 Breast, pem_pos=1, final_scores=-1.209234, final_local_scores=-1.2122369, pem=-0.3169917, similarity=-30.362003]

gt: 70547 Breast_cancer, pem_pos=0, final_scores=-1.3394594, final_local_scores=-1.3925183, pem=0.0, similarity=-37.465813]

[winner: 8263081 Nasopharynx_cancer, pem_pos=5, final_scores=-1.2926874, final_local_scores=-1.2787406, pem=-1.609438, similarity=-25.31808]

gt: 70547 Breast_cancer, pem_pos=10, final_scores=-2.532873, final_local_scores=-2.696141, pem=-6.7711987, similarity=-39.545677]

3: 146315 Bleeding

[winner: 146315 Bleeding, pem_pos=0, final_scores=-1.0867426, final_local_scores=-1.072852, pem=-0.036367625, similarity=-27.72856]

gt: 146315 Bleeding, pem_pos=0, final_scores=-1.0867426, final_local_scores=-1.072852, pem=-0.036367625, similarity=-27.72856]

4: 1172094 Radiosurgery

[winner: 1172094 Radiosurgery, pem_pos=0, final_scores=-0.91309065, final_local_scores=-0.91377044, pem=0.0, similarity=-23.17902]

gt: 1172094 Radiosurgery, pem_pos=0, final_scores=-0.91309065, final_local_scores=-0.91377044, pem=0.0, similarity=-23.17902]

5: 105219 Cancer

[winner: 105219 Cancer, pem_pos=0,

```

final_scores=-1.502543, final_local_scores=-1.5863984,
pem=-0.060834263, similarity=-42.920753|
    gt: 105219 Cancer, pem_pos=0, final_scores=-1.502543,
final_local_scores=-1.5863984, pem=-0.060834263,
similarity=-42.920753 ]
6: 103915 Tissue_(biology)
    [winner: 103915 Tissue_(biology), pem_pos=0,
final_scores=-1.0802547, final_local_scores=-1.0789154,
pem=0.0, similarity=-28.107271|
    gt: 103915 Tissue_(biology), pem_pos=0,
final_scores=-1.0802547, final_local_scores=-1.0789154,
pem=0.0, similarity=-28.107271 ]
7: 3054853 Three-dimensional_space
    [winner: 38267 Dimension_(vector_space), pem_pos=1,
final_scores=-0.9957579, final_local_scores=-0.9218852,
pem=-0.51930404, similarity=-20.597124|
    gt: 3054853 Three-dimensional_space, pem_pos=10,
final_scores=-2.525197, final_local_scores=-2.678734,
pem=-7.9455557, similarity=-32.639885 ]
    [winner: 374790 Germantown, New York, pem_pos=3,
final_scores=-1.1931725, final_local_scores=-1.1019204,
pem=-2.533128, similarity=-15.018241|
    RECALL MISS: 3054853 Three-dimensional_space]
8: 1189445 Stereotactic_surgery
    [winner: 9300815 Synovectomy, pem_pos=3,
final_scores=-0.8575518, final_local_scores=-0.7334432,
pem=-1.789957, similarity=-8.063631|
    RECALL MISS: 1189445 Stereotactic_surgery]
9: 25856 Radiation
    [winner: 25856 Radiation, pem_pos=0,
final_scores=-1.3773987, final_local_scores=-1.4434812,
pem=-0.48442143, similarity=-36.352287|
    gt: 25856 Radiation, pem_pos=0, final_scores=-1.3773987,
final_local_scores=-1.4434812, pem=-0.48442143,
similarity=-36.352287 ]
10: 273524 Linear_particle_accelerator
    [winner: 273524 Linear_particle_accelerator, pem_pos=0,
final_scores=-0.1268608, final_local_scores=0.03851983,
pem=0.0, similarity=5.2392216|
    gt: 273524 Linear_particle_accelerator, pem_pos=0,
final_scores=-0.1268608, final_local_scores=0.03851983,
pem=0.0, similarity=5.2392216 ]
11: 5664 Consciousness
    [winner: 5664 Consciousness, pem_pos=1,
final_scores=-1.3495594, final_local_scores=-1.3639543,
pem=-0.07470379, similarity=-36.207157|
    gt: 5664 Consciousness, pem_pos=1,
final_scores=-1.3495594, final_local_scores=-1.3639543,
pem=-0.07470379, similarity=-36.207157 ]
12: 321263 Congenital_disorder
    [winner: 321263 Congenital_disorder, pem_pos=0,
final_scores=-1.3259518, final_local_scores=-1.3528411,

```

```
pem=0.0, similarity=-36.281765|
    gt: 321263 Congenital_disorder, pem_pos=0,
final_scores=-1.3259518, final_local_scores=-1.3528411,
pem=0.0, similarity=-36.281765 ]
13: 212193 Coagulation
    [winner: 212193 Coagulation, pem_pos=0,
final_scores=-1.1799226, final_local_scores=-1.178766,
pem=-0.2845832, similarity=-29.539406|
    gt: 212193 Coagulation, pem_pos=0,
final_scores=-1.1799226, final_local_scores=-1.178766,
pem=-0.2845832, similarity=-29.539406 ]
14: 56512 Capillary
    [winner: 56512 Capillary, pem_pos=0,
final_scores=-1.06552, final_local_scores=-1.0219266,
pem=0.0, similarity=-26.406614|
    gt: 56512 Capillary, pem_pos=0, final_scores=-1.06552,
final_local_scores=-1.0219266, pem=0.0,
similarity=-26.406614 ]
1: 56615 Bangkok
    [pem_pos=0, final_scores=0.48701274,
final_local_scores=0.48847407, pem=-0.0005963434,
similarity=18.669996]
2: 30128 Thailand
    [pem_pos=0, final_scores=0.42215496,
final_local_scores=0.3856659, pem=-0.24029273,
similarity=16.9055]
1: 30128 Thailand
    [pem_pos=0, final_scores=0.30290025,
final_local_scores=0.25539824, pem=-0.24029273,
similarity=13.018055]
```